

BEST AVAILABLE COPY



PCT/AU99/00385

out

REC'D 29 JUN 1999	
WIPO	PCT

EPO - DG 1

Patent Office
Canberra

20.12.2000

(101)

I, KIM MARSHALL, MANAGER EXAMINATION SUPPORT AND SALES,
hereby certify that the annexed is a true copy of the Provisional specification in
connection with Application No. PP 3634 for a patent by THE UNIVERSITY OF
SYDNEY filed on 21 May 1998.



WITNESS my hand this Seventeenth
day of June 1999

KIM MARSHALL
MANAGER EXAMINATION SUPPORT AND
SALES

**PRIORITY
DOCUMENT**

SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

AUSTRALIA
Patents Act 1990

PROVISIONAL SPECIFICATION

Applicants:

THE UNIVERSITY OF SYDNEY

Invention Title:

ANTIGENS AND THEIR DETECTION

The invention is described in the following statement:

Antigens and Their Detection

TECHNICAL FIELD

The invention relates to novel nucleotide sequences located in a gene which encodes a bacterial flagellin antigen, and the use of those nucleotide sequences for the detection of bacteria which express particular flagellin antigens, on the basis of that antigen alone, or in conjunction with the O antigen expressed by that strain.

BACKGROUND ART

The flagellum of many bacteria appears to be made up of a single protein known as flagellin. The serotyping schemes of *E. coli* and *Salmonella enterica* are based on highly variable antigenic surface structures which include the lipopolysaccharide which carries the O antigen and flagellin which is now known to be the carrier of the classical H antigen. In many strains of *S. enterica* there are two loci (*fliC* and *fljB*) which encode flagellin, and a regulatory system which allows one only to be expressed at any time; and which also provides for expression to rapidly alternate between the two forms first identified as two phases (H1 and H2) for the H antigen of most strains. In *E. coli* there are 54 forms of H antigen recognised and until recently they were all thought to be encoded at the *fliC* locus, as has been shown for *E. coli* K-12. However in the 1980s Ratiner [Ratiner Y A "Phase variation of the H antigen in *Escherichia coli* strain Bi327-41, the standard strain for *Escherichia coli* flagellin antigen H3" FEMS Microbiol. Lett 15 (1982) 33-36; Ratiner Y A "Presence of two structural genes determining antigenically different phase-specific flagellins in some *Escherichia coli* strains" FEMS Microbiol. Lett. 19 (1983) 37-41; Ratiner Y A "Two genetic arrangements determining flagellin antigen

specificities in two diphasic *Escherichia coli* strains" FEMS Microbiol. Lett. 29 (1985) 317-323; Ratiner Y A "Different alleles of the flagellin gene *hagB* in *Escherichia coli* standard H test strains" FEMS Microbiol Lett. 48 (1987) 97-104.] showed that in some cases there are two loci and that expression can alternate. The matter was further complicated by a recent paper by Ratiner [Ratiner Y A (1998) "New flagellin-specifying genes in some *Escherichia coli* strains" J. Bacteriol. 180 979-984] showing three loci (*flk*, *fll* and *flm*) for flagellin in addition to *fliC* although the *fljB* locus has not been found in *E. coli*. However *E. coli* strains are normally identified by the combination of one O antigen and one H antigen [and K antigen when present as a capsule (K) antigen], with no problems reported for the vast majority of cases with alternate phases, while *S. enterica* strains are normally identified by the combination of O, H1 and H2 antigens. It is still not clear how widespread in *E. coli* H antigens determined by flagellin genes other than *fliC* are.

Typing is typically carried out using specific antisera. The incidence of pathogenic *E. coli* in association with human and animal disease supports the need for suitable and rapid typing techniques.

DESCRIPTION OF THE INVENTION

In a first aspect, the present invention provides a novel nucleic acid molecule encoding all or part of an *E. coli* flagellin protein.

The present invention provides, for the first time, full length sequence for a flagellin gene for the following *E. coli* serotypes: H6, H9, H10, H14, H18, H23, H51, H45, H49, H19, H30, H32, H26, H41, H15, H16, H20, H28, H46, H31, H34, H43 and H52. Corrected full length sequences have been obtained for H7 and H12.

Partial flagellin gene sequence, including the central variable region, has been obtained for the following *E. coli* H serotypes: H40, H8, H21, H47, H11, H17, H25, H42, H27, H35, H2, H3, H24, H37, H50, H4, H44, H38, H39, H55, H29, H33, H5, H54 and H56. Comparison of sequences demonstrates that unique flagellin genes have now been sequenced (partially or completely) for the following *E. coli* H serotypes: H2, H3, H5, H6, H7, H9, H11, H14, H18, H19, H20, H21, H23, H24, H25, H26, H27, H28, H29, H30, H31, H32, H33, H34, H35, H37, H38, H39, H41, H42, H43, H45, H46, H48, H49, H51, H52, H54, and H56 and either H8 or H40, H15 or H16, H1 or H12, H10 or H50 and H4 or H17.

By comparison of these sequences, the present inventors were able to identify specific sequences for each of the above H serotypes.

The present invention also provides *fliC* sequences from 10 different H7 strains, in addition to that from the H7 typing strain, and two sequences specific to H7 of O157 and O55 *E. coli* strains.

The present invention encompasses all or part of the unique genes sequenced for H2, H3, H5, H6, H9, H11, H14, H18, H19, H20, H21, H23, H24, H25, H26, H27, H28, H29, H30, H31, H32, H33, H34, H35, H37, H38, H39, H41, H42, H43, H45, H46, H48, H49, H51, H52, H54 and H56 and either H8 or H40, H15 or H16, H10 or H50 and H4 or H17. The invention also encompasses newly provided sequence for H7 and H12 as well as novel primers for the specific amplification of H1, H7, H12 and H48 as well as for the other above mentioned newly sequenced flagellin genes.

The nucleic acid molecules of the invention may be variable in length. In one embodiment they are oligonucleotides of from about 10 to about 20 nucleotides in length. The oligonucleotides of the invention are specific for the flagellin gene from which they are derived and are derived from the central region of the

gene. In one embodiment, oligonucleotides in accordance with the present invention, which also include oligonucleotides from the previously sequenced *E. coli* H1, H7, H12 and H48 genes, are those shown in Table 3.

5 The 44 sequences (see Table 3) provide a panel to which newly sequenced genes can be compared to select specific oligonucleotides for those newly sequenced genes.

 In a second aspect the invention provides a method of detecting the presence of *E. coli* of a particular H
10 serotype in a sample, the method comprising the step of specifically hybridising at least one nucleic acid molecule derived from a flagellin gene, wherein the at least one nucleic acid molecule is specific for a particular flagellin gene associated with the H serotype,
15 to any *E. coli* in the sample which contain the gene, and detecting any specifically hybridised nucleic acid molecules, wherein the presence of specifically hybridised nucleic acid molecules identifies the presence of the H serotype in the sample.

20 In one preferred embodiment the detection method is a Southern blot method. More preferably, the nucleic acid molecule is labelled and hybridisation of the nucleic acid molecule is detected by autoradiography or detection of fluorescence.

25 Preferred nucleic acid molecules for the detection of particular flagellin genes are listed in Table 3.

 In a third aspect the invention provides a method of detecting the presence of *E. coli* of a particular H
30 serotype in a sample, the method comprising the step of specifically hybridising at least one pair of nucleic acid molecules to any *E. coli* in the sample which contains the flagellin gene for the particular H serotype, wherein at least one of the nucleic acid molecules is specific for the particular flagellin gene associated with the H
35 serotype, and detecting any specifically hybridised

nucleic acid molecules, wherein the presence of specifically hybridised nucleic acid molecules identifies the presence of the H serotype in the sample.

5 In one preferred embodiment the detection method is a polymerase chain reaction method. More preferably, the nucleic acid molecules are labelled and hybridisation of the nucleic acid molecule is detected by electrophoresis.

10 It is recognised that there may be instances where spurious hybridisation will arise through the initial selection of a sequence found in many different genes but this is typically recognisable by, for instance, comparison of band sizes against controls in PCR gels, and an alternative sequence can be selected.

15 In a fourth aspect the invention provides a method for detecting the presence of a particular O serotype and H serotype of *E. coli* in a sample, the method comprising the following steps:

20 (a) specifically hybridising at least one nucleic acid molecule, derived from and specific for a gene encoding a transferase or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit, the gene being involved in the synthesis of a particular *E. coli* O antigen, to any *E. coli* in the sample which contain the gene;

25 (b) specifically hybridising at least one nucleic acid molecule derived from and specific for a particular flagellin gene associated with that H serotype, to any *E. coli* in the sample which contain the gene; and

30 (c) detecting any specifically hybridised nucleic acid molecules.

Preferred nucleic acid molecules for the detection of particular flagellin genes are listed in Table 3.

35 In one preferred embodiment, the sequence of the nucleic acid molecule specific for the O antigen is specific to the nucleotide sequence encoding the O111

antigen. More preferably, the sequence is derived from a gene selected from the group consisting of *wbdH* (nucleotide position 739 to 1932 of Figure 5), *wzx* (nucleotide position 8646 to 9911 of Figure 5), *wzy* (nucleotide position 9901 to 10953 of Figure 5), *wbdM* (nucleotide position 11821 to 12945 of Figure 5) and fragments of those molecules of at least 10-12 nucleotides in length. Particularly preferred nucleic acid molecules are those set out in Tables 8 and 8A, with respect to the above mentioned genes.

In another preferred embodiment, the sequence of the nucleic acid molecule specific for the O antigen is specific to the nucleotide sequence encoding the O157 antigen. More preferably, the sequence is derived from a gene selected from the group consisting of *wbdN* (nucleotide position 79 to 861 of Figure 6), *wbdO* (nucleotide position 2011 to 2757 of Figure 6), *wbdP* (nucleotide position 5257 to 6471 of Figure 6), *wbdR* (nucleotide position 13156 to 13821 of Figure 6), *wzx* (nucleotide position 2744 to 4135 of Figure 6) and *wzy* (nucleotide position 858 to 2042 of Figure 6) and fragments of those molecules of at least 10-12 nucleotides in length. Particularly preferred nucleic acid molecules are those set out in Tables 9 and 9A, with respect to the above mentioned genes.

In one preferred embodiment the detection method is a Southern blot method. More preferably, the nucleic acid molecule is labelled and hybridisation of the nucleic acid molecule is detected by autoradiography or detection of fluorescence.

In a fifth aspect the invention provides a method for detecting the presence of a particular O serotype and H serotype of *E. coli* in a sample, the method comprising the following steps:

(a) specifically hybridising at least one pair of

nucleic acid molecules, at least one of which is derived from and specific for a gene encoding a transferase or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit, the gene being involved in the synthesis of the particular *E. coli* O antigen, to any *E. coli* in the sample which contain the gene;

(b) specifically hybridising at least one pair of nucleic acid molecules, at least one of which is derived from and specific for a particular flagellin gene associated with the particular H serotype, to any *E. coli* in the sample which contain the gene; and

(c) detecting any specifically hybridised nucleic acid molecules.

Preferred nucleic acid molecules for the detection of particular flagellin genes are listed in Table 3.

In one preferred embodiment, the sequence of the nucleic acid molecule specific for the O antigen is specific to the nucleotide sequence encoding the O111 antigen. More preferably, the sequence is derived from a gene selected from the group consisting of *wbdH* (nucleotide position 739 to 1932 of Figure 5), *wzx* (nucleotide position 8646 to 9911 of Figure 5), *wzy* (nucleotide position 9901 to 10953 of Figure 5), *wbdM* (nucleotide position 11821 to 12945 of Figure 5) and fragments of those molecules of at least 10-12 nucleotides in length. Particularly preferred nucleic acid molecules are those set out in Tables 8 and 8A, with respect to the above mentioned genes.

In another preferred embodiment, the sequence of the nucleic acid molecule specific for the O antigen is specific to the nucleotide sequence encoding the O157 antigen. More preferably, the sequence is derived from a gene selected from the group consisting of *wbdN* (nucleotide position 79 to 861 of Figure 6), *wbdO* (nucleotide position

2011 to 2757 of Figure 6), *wbdP* (nucleotide position 5257 to 6471 of Figure 6), *wbdR* (nucleotide position 13156 to 13821 of Figure 6), *wzx* (nucleotide position 2744 to 4135 of Figure 6) and *wzy* (nucleotide position 858 to 2042 of Figure 6) and fragments of those molecules of at least 10-12 nucleotides in length. Particularly preferred nucleic acid molecules are those set out in Tables 9 and 9A, with respect to the above mentioned genes.

In one preferred embodiment the detection method is a polymerase chain reaction method. More preferably, the nucleic acid molecules are labelled and hybridisation of the nucleic acid molecule is detected by electrophoresis.

The present inventors believe that based on the teachings of the present invention and available information concerning O antigen gene clusters, and through use of experimental analysis, comparison of nucleic acid sequences or predicted protein structures, nucleic acid molecules in accordance with the invention can be readily derived for any particular O antigen of interest. Suitable bacterial strains can typically be acquired commercially from depositary institutions.

There are currently 166 defined *E. coli* O antigens.

Samples of the 166 different *E. coli* O antigen serotypes are available from Statens Serum Institut, Copenhagen, Denmark.

The inventors envisage rare circumstances whereby two genetically similar gene clusters encoding serologically different O antigens have arisen through recombination of genes or mutation so as to generate polymorphic variants. In these circumstances multiple pairs of oligonucleotides may be selected to provide hybridisation to the specific combination of genes. The invention thus envisages the use of a panel containing multiple nucleic acid molecules for use in the method of testing for O antigen in conjunction with H antigen, wherein the nucleic acid

molecules are derived from genes encoding transferases and/or enzymes for the transport or processing of a polysaccharide or oligosaccharide unit including wzx or wzy genes, wherein the panel of nucleic acid molecules is specific to a particular O antigen. The panel of nucleic acid molecules can include nucleic acid molecules derived from O antigen sugar pathway genes where necessary.

The inventors also found two mutated flagellin genes from H typing strains for H35 and H54 which have insertion sequences inserted into normal flagellar genes identical or near identical to that of the H11 and H21 typing strains respectively. Thus, primers for H11 and H21 (listed in Table 3) would also amplify fragments in H35 and H54, which differ in sizes to those in H11 and H21 respectively. The inventors also provide two pairs of primers each for H35 and H54 based on the insertion sequence (see H35 and H54 columns in Table 3). The use of one of them in combination with one of the H11 or H21 primers will generate a PCR band only in H35 or H54 respectively, and this will also differentiate H35 and H54 from H11 and H21 respectively.

The present invention also relates to methods of detecting the presence of particular *E. coli* H antigens or H antigen and O antigen combinations where one or more nucleic acid molecules which generate a particular size fragment indicative of the presence of that H antigen are used or in which the combination of one antigen specific primer for that H antigen with another primer for a related H antigen provides for the detection of the particular H antigen by hybridisation to the relevant gene. Preferably, the H antigen is H11, H21, H35 or H54.

The pairs of nucleic acid molecules where the method of the fifth aspect is used may both hybridise to the relevant H or O antigen gene or alternatively only one may hybridise to the relevant gene and the other to another

site.

The inventors recognise in applying the methods of the invention for detecting combinations of O and H antigens to samples, that the methods do not indicate whether a positive result for a particular O and H antigen combination arises because the O and H antigen are present on a single *E. coli* strain present in the sample or are present on different *E. coli* strains present in the sample. Because the ability to identify the presence of *E. coli* strains with particular O and H antigen combinations is highly desirable (due to the relationship between particular combinations and pathogenicity) the determination that a particular combination is present in a sample can be followed by isolation of single colonies and checking whether they contain the relevant combination by using the same method again or using antibody labelled magnetic beads to separate cells expressing the particular O or H antigen and then testing the isolated cells for the other serotype.

In addition, as mentioned above, the present inventors have established the existence of H7 primers specific to the O157 and O55 serotypes. Using such primers it is possible to detect particular O and H antigen combinations with the use of H specific nucleic acid molecules.

In a sixth aspect the invention provides a method for detecting the presence of a particular O serotype and H serotype of *E. coli* in a sample, the method comprising the following steps:

(a) specifically hybridising at least one nucleic acid molecule, derived from and specific for a gene encoding a flagellin associated with a particular *E. coli* H antigen serotype to any *E. coli* carrying the gene and present in the sample;

and

(b) detecting the at least one specifically hybridised nucleic acid molecule, wherein the at least one nucleic acid molecule is specific for the particular combination of O and H antigen.

Preferably the combination is O55:H7 or O157:H7.

The ability to detect the O157:H7 combination from a particular H7 primer or pair is of particular use given the association of this combination with pathogenic strains.

In a seventh aspect the present invention provides a method for testing a food derived sample for the presence of one or more particular *E. coli* O antigens and H antigens comprising testing the sample by a method of the fourth, fifth or sixth aspect the invention.

In an eighth aspect the present invention provides a method for testing a faecal derived sample for the presence of one or more particular *E. coli* O antigens and H antigens comprising testing the sample by a method of the fourth, fifth or sixth aspect the invention.

In a ninth aspect the present invention provides a method for testing a patient or animal derived sample for the presence of one or more particular *E. coli* O antigens and H antigens comprising testing the sample by a method of the fourth, fifth or sixth aspect the invention.

Preferably, the method of the seventh, eighth or ninth aspect of the invention is a polymerase chain reaction method. More preferably the oligonucleotide molecules for use in the method are labelled. Even more preferably the hybridised nucleic acid molecules are detected by electrophoresis.

In the above described methods it will be understood that where pairs of nucleic acid molecules are used one of the nucleic acid molecules may hybridise to a sequence that is not from the O antigen transferase, *wzx* or *wzy*

gene or the flagellin gene. Further where both hybridise to these genes the O antigen molecules may hybridise to the same or a different one of these genes.

In a tenth aspect the present invention provides a kit for identifying the H serotype of *E. coli*, the kit comprising:

at least one nucleic acid molecule derived from and specific for an *E. coli* flagellin gene.

In an eleventh aspect the present invention provides a kit for identifying the H and O serotype of *E. coli*, the kit comprising:

(a) at least one nucleic acid molecule derived from and specific for an *E. coli* flagellin gene; and

(b) at least one nucleic acid molecule derived from and specific for a gene encoding a transferase or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit, the gene being involved in the synthesis of a particular *E. coli* O antigen.

The nucleic acid molecules may be provided in the same or different vials. The kit may also provide in the same or separate vials a second set of specific nucleic acid molecules.

Particularly preferred nucleic acid molecules for inclusion in the kits are those specified in Tables 3, 8, 8A, 9 and 9A as described above.

DEFINITIONS

In this specification, we have used term "flagellin gene" in many cases where previously one would have used "*fliC*", to allow for the uncertainty as to locus introduced by recent observations. However, uncertainty as to the locus does not alter the fact that most *E. coli* strains express a single H antigen and that a single flagellin gene sequence

per strain is required to give the genetic basis for H antigen variation . Any use of the name *fliC* in this specification where a different locus is later shown to be involved would not affect the validity of conclusions drawn regarding application of information based on the sequence, where the conclusions do not relate to the map position. Thus it is generally the nucleic acid molecule itself which is of importance rather than the name attributed to the gene. When it is known or suspected that the gene encoding the H antigen is not in the *fliC* locus, we use the term flagellin rather than *fliC*.

The phrase, "a nucleic acid molecule derived from a gene" means that the nucleic acid molecule has a nucleotide sequence which is either identical or substantially similar to all or part of the identified gene. Thus a nucleic acid molecule derived from a gene can be a molecule which is isolated from the identified gene by physical separation from that gene, or a molecule which is artificially synthesised and has a nucleotide sequence which is either identical to or substantially similar to all or part of the identified gene. While some workers consider only the DNA strand with the same sequence as the mRNA transcribed from the gene, here either strand is intended.

Transferase genes are regions of nucleic acid which have a nucleotide sequence which encodes gene products that transfer monomeric sugar units.

Flippase or *wzx* genes are regions of nucleic acid which have a nucleotide sequence which encodes a gene product that flips oligosaccharide repeat units generally composed of three to six monomeric sugar units to the external surface of the membrane.

Polymerase or *wzy* genes are regions of nucleic acid which have a nucleotide sequence which encodes gene products that polymerise repeating oligosaccharide units

generally composed of 3-6 monomeric sugar units.

The nucleotide sequences provided in this specification are described as anti-sense sequences. This term is used in the same manner as it is used in Glossary of Biochemistry and Molecular Biology Revised Edition, David M. Glick, 1997 Portland Press Ltd., London on page 11 where the term is described as referring to one of the two strands of double-stranded DNA usually that which has the same sequence as the mRNA. We use it to describe this strand which has the same sequence as the mRNA.

NOMENCLATURE

Synonyms for E. coli O111 *rfb*

	<u>Current names</u>	<u>Our names</u>	<u>Bastin et al. 1991</u>
15	wbdH	orf1	
	gmd	orf2	
	wbdI	orf3	orf3.4*
	manC	orf4	rfbM*
	manB	orf5	rfbK*
20	wbdJ	orf6	orf6.7*
	wbdK	orf7	orf7.7*
	wzx	orf8	orf8.9 and rfbX*
	wzy	orf9	
	wbdL	orf10	
25	wbdM	orf11	

* Nomenclature according to Bastin D.A., et al. 1991 "Molecular cloning and expression in Escherichia coli K-12 of the *rfb* gene cluster determining the O antigen of an E. coli O111 strain". *Mol. Microbiol.* 5:9 2223-2231.

Other Synonyms

	wzy	rbc
	wzx	rfbX
	rmlA	rfbA
35	rmlB	rfbB
	rmlC	rfbC
	rmlD	rfbD
	glf	orf6*
	wbbI	orf3#, orf8* of <u>E. coli</u> K-12
40	wbbJ	orf2#, orf9* of <u>E. coli</u> K-12
	wbbK	orf1#, orf10* of <u>E. coli</u> K-12
	wbbL	orf5#, orf 11* of <u>E. coli</u> K-12
	#	Nomenclature according to Yao, Z. And M. A. Valvano 1994.

"Genetic analysis of the O-specific lipopolysaccharide biosynthesis region (rfb) of Escherichia coli K-12 W3110: identification of genes the confer groups-specificity to Shigella flexneri serotypes Y and 4a". J. Bacteriol. 176: 4133-4143.

5 * Nomenclature according to Stevenson et al. 1994. "Structure of the O-antigen of E. coli K-12 and the sequence of its rfb gene cluster". J. Bacteriol 176: 4144-4156.

10 • The O antigen genes of many species were given rfb names (rfbA etc) and the O antigen gene cluster was often referred to as the rfb cluster. There are now new names for the rfb genes as shown in the table. Both terminologies have been used herein, depending on the source of the information.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows *Eco* R1 restriction maps of cosmid clones pPR1054, pPR1055, pPR1056, pPR1058, pPR1287 which are subclones of *E. coli* O111 O antigen gene cluster. The thickened line is the region common to all clones. Broken lines show segments that are non-contiguous on the chromosome. The deduced restriction map for *E. coli* strain M92 is shown above.

Figure 2 shows a restriction mapping analysis of *E. coli* O111 O antigen gene cluster within the cosmid clone pPR1058. Restriction enzymes are: (B: *Bam*HI; Bg: *Bgl*II, E: *Eco*R1; H: *Hind*III; K: *Kpn*I; P: *Pst*I; S: *Sal*I and X: *Xho*I. Plasmids pPR1230, pPR1231, and pPR1288 are deletion derivatives of pPR1058. Plasmids pPR 1237, pPR1238, pPR1239 and pPR1240 are in pUC19. Plasmids pPR1243, pPR1244, pPR1245, pPR1246 and pPR1248 are in pUC18, and pPR1292 is in pUC19. Plasmid pPR1270 is in pT7T319U. Probes 1, 2 and 3 were isolated as internal fragments of pPR1246, pPR1243 and pPR1237 respectively. Dotted lines indicate that subclone DNA extends to the left of the map into attached vector.

Figure 3 shows the structure of *E. coli* O111 O antigen gene cluster.

Figure 4 shows the structure of *E. coli* O157 O antigen gene cluster.

Figure 5 shows the nucleotide sequence of the *E. coli* O111 O antigen gene cluster. Note: (1) The first and last three bases of a gene are underlined and of italic respectively.; (2) The region which was previously sequenced by Bastin and Reeves 1995 "Sequence and analysis of the O antigen gene (*rfb*) cluster of *Escherichia coli* O111" Gene 164: 17-23 is marked.

Figure 6 shows the nucleotide sequence of the *E. coli* O157 O antigen gene cluster. Note: (1) The first and last

three bases of a gene (region) are underlined and of *italic* respectively (2) The region previously sequenced by Bilge et al. 1996 "Role of the *Escherichia coli* O157-H7 O side chain in adherence and analysis of an rfb locus". Inf. and Immun 64:4795-4801 is marked.

Figures 7 to 18 show the nucleotide sequences obtained for flagellin genes from *E. coli* typing strains for H1-H12 respectively. The primer positions listed in Table 3 are based on treating the first nucleotide of each of these sequences as No. 1.

Figures 19 to 26 show the nucleotide sequences obtained for flagellin genes from *E. coli* typing strains for H14-H21 respectively. The primer positions listed in Table 3 are based on treating the first nucleotide of each of these sequences as No. 1.

Figures 27 to 39 show the nucleotide sequences obtained for flagellin genes from *E. coli* typing strains for H23-H35 respectively. The primer positions listed in Table 3 are based on treating the first nucleotide of each of these sequences as No. 1.

Figures 40 to 55 show the nucleotide sequences obtained for flagellin genes from *E. coli* typing strains for H37-H52 respectively. The primer positions listed in Table 3 are based on treating the first nucleotide of each of these sequences as No. 1.

Figures 56 to 58 show the nucleotide sequences obtained for flagellin genes from *E. coli* typing strains for H54-H56 respectively. The primer positions listed in Table 3 are based on treating the first nucleotide of each of these sequences as No. 1.

Figures 59 to 68 show the nucleotide sequences obtained for flagellin genes from *E. coli* H7 strains M1179, M1004, M1211, M1200, M1686, M1328, M917, M527, M973 and M918 respectively. The primer positions listed in Table 3 are based on treating the first nucleotide of each of these

sequences as No. 1.

BEST METHOD OF CARRYING OUT THE INVENTION

5 In carrying out the methods of the invention with respect
to the testing of particular sample types including samples
from food, patients, animals and faeces the samples are
prepared by routine techniques routinely used in the
preparation of such samples for DNA based testing. The
10 steps for testing the samples using particular nucleic acid
molecules in assay formats such as Southern blots and PCR
are performed under routinely determined conditions
appropriate to the sample and the nucleic acid molecules.

15 H antigen

Materials and Methods

1. Bacterial strains:

There are 54 H types in *E. coli* [Ewing, W.H.: Edwards and
Ewing's identification of the *Enterobacteriaceae*, Elsevier
20 Science Publishers, Amsterdam, The Netherlands, 1986]: note
H antigens from 1 to 57 were listed and that 13, 22 and 57
are not valid. The standard H type strains were obtained
from the Institute of Medical and Veterinary Science,
Adelaide, Australia. The primary stocks are hold at the
25 Statens Serum Institut, Copenhagen, Denmark.

The additional H7 strains used are listed in Table 1.

2. Isolation of chromosomal DNA:

Chromosomal DNA from all the 54 H type strains and the
30 strains listed in Table 1 was isolated using the Promega
Genomic isolation kit (Madison WI USA). Each chromosomal
DNA sample was checked by gel electrophoresis of the DNA
and by PCR amplification of the *mdh* gene using
oligonucleotides based on the *E. coli* K-12 *mdh* gene [Boyd,
35 E.F., Nelson, K., Wang, F.-S., Whittam, T.S. and Selander,

R.K.: Molecular genetic basis of allelic polymorphism in malate dehydrogenase (*mdh*) in natural populations of *Escherichia coli* and *Salmonella enterica*. Proc. Natl. Acad. Sci. USA 91 (1994) 1280-1284].

5

3. PCR amplification of flagellin gene:

Flagellin genes from different strains were first PCR amplified using one of the following four pairs of oligonucleotides:

10

#1285 (5'-atggcacaagtcattaatac) and

#1286 (5'-ttaaccctgcagtagagaca);

#1417 (5'-ctgatcactcaaaataatatcaac) and

#1418 (5'-ctgcggtacctggttggc);

15

#1431 (5'-atggcacaagtcattaatacccaac) and

#1432 (5'-ctaaccctgcagcagagaca):

#1575 (5'-gggtggaaaccaatac) and

#1576 (5'-gcgcacaggcaatttgg)

PCR reactions were carried out under the following

20

conditions: denaturing, 94°C/30'; annealing, temperature varies (refer to Table 2)/30'; extension, 72°C/1'; 30 cycles. The PCR product was purified using the Promega Wizard PCR purification kit (Madison WI USA) before being sequenced.

25

The H36 and H53 strains gave two PCR bands using primer pairs #1431/#1432 and #1417/#1418 respectively, and were not sequenced.

4. Sequencing of the flagellin genes:

30

Each PCR product was first sequenced using the oligonucleotide primers used for the PCR amplification. Primers based on the obtained sequence were then used to sequence further, and this procedure was repeated until the entire PCR product was sequenced.

The sequencing reactions were performed using the DyeDeoxy Terminator Cycle Sequencing method (Applied Biosystems, CA, USA), and reaction products were analysed using fluorescent dye and an ABI377 automated sequencer (CA, USA).

Sequence data were processed and analysed using Staden programs [Sacchi CT, Zanella R C, Caugant D A, Frasch C E, Hidalgo N T, Milagres L G, Pessoa L L, Ramos S R, Camargo M C C and Melles C E A "Emergence of a new clone of serogroup *C Neisseria meningitidis* in Sao Paulo, Brazil" J. Clin. Microbiol. 30 (1992) 1282-1286;

Staden, R.: Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. Nucl. Acids Res. 10 (1982a) 4731-4751;

Staden, R.: An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. Nucl. Acids Res. 10 (1982b) 2951-2961;

Staden, R.: Computer methods to locate signals in nucleic acid sequences. Nucl. Acids Res. 12 (1984a) 505-519;

Staden, R.: Graphic methods to determine the function of nucleic acid sequences. A summary of ANALYSEQ options. Nucl. Acids Res. 12 (1984b) 521-538;

Staden, R.: The current status and portability of our sequence handling software. Nucl. Acids Res. 14 (1986) 217-231].

We were able to PCR amplify flagellin genes from H typing strains for H7, 23, 12, 51, 45, 49, 19, 9, 30, 32, 26, 41, 15, 16, 20, 28, 46, 31, 14, 18, 6, 34, 48, 43, 10, 52, and also from H7 strains m1004, m527, m1686, m1211, m1328, m973, m1179, m1200, m917, and m918 using primers #1575 and #1576 which are based on sequences 51-34 bp upstream and 37-54 bp downstream of start and end of the *E. coli* K-12 *fliC* gene respectively. Thus, the full sequence of the flagellin gene from these strains was obtained and the use of flanking sequence for primers makes it highly

likely that they are at the *fliC* locus.

For other strains, we were only able to amplify the *flagellin* gene using one or more of the other three pairs of primers, which are based on sequence within the *fliC* gene, and thus only partial sequence was obtained. These amplicons may be of the *fliC* gene or one of the alternative flagellin genes. The flagellin gene sequences from H typing strains for H40, 8, 21, 47, 11, 27, 35, 2, 3, 24, 37, 50, 4, 44, 38, 55, 29, 33, 5, and 56 obtained are lacking 18 and 14 codons at 5' and 3' ends respectively. The flagellin gene sequence of H39 obtained using primers #1285/#1286 lacks 18 and 19 codons at 5' and 3' ends respectively. The flagellin gene sequence of H typing strains of H17, 25 and 42 lack 23 and 21 codons at 5' and 3' ends respectively. The flagellin gene sequence of the H typing strain for H54 lacks 23 and 12 codons at the 5' and 3' ends respectively. There is very little variation in the sequence at the two ends of flagellin genes and antigenic variation is due to variation in the central region of the gene. The absence of sequence for the ends of some of the flagellin genes is not important for the purpose of the present invention relating to the detection of antigenic variation by DNA sequence based means.

The *fliC* genes from H type strains of H1, H7 and H12 have been sequenced previously [Schoenhals, G. and Whitfield, C.: Comparative analysis of flagellin sequences from *Escherichia coli* strains possessing serologically distinct flagellar filaments with a shared complex surface pattern. J. Bacteriol. 175 (1993) 5395-5402] and we did not sequence the gene from the H1 strain.

We have sequenced *fliC* genes from a set of H7 strains with different O antigens, including that of *fliC* from the H7 typing strain as one of the set: we have found four differences from the published H7 sequence (GenBank accession number L07388) which we believe are due to errors

in the published sequence.

We have also re-sequenced the *fliC* gene from the H12 type strain, and have found one difference from the published H12 sequence (GenBank accession number L07389) which we believe is due to an error in the published sequence.

The flagellin genes from type strains H35 and H54 were also amplified using primers #1431/#1432, which are based on sequence within the *fliC* gene. Sequence data revealed that these two genes would be non-functional due to insertion sequence inserted in the middle of them. We have sequenced them to facilitate selection of primers for the functional flagellin genes.

5. Comparison and alignment of the flagellin genes:

Programs Pileup [Devereux, J., Haeberli, P. and Smithies, O.: A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. 12 (1984) 387-395] and Multicomp [Reeves, P.R., Farnell, L. and Lan, R.: MULTICOMP: a program for preparing sequence data for phylogenetic analysis. CABIOS 10 (1994) 281-284] were used.

The previously published sequence of H1 (GenBank accession number L07387) was extracted from GenBank and used. Because we did not sequence H36 and H53 flagellin genes, we only compared 52 flagellin genes of H typing strains and the *fliC* genes from the additional 10 H7 strains.

Among the H7 *fliC* genes, the percentage of DNA difference ranged from 0.0 to 2.39%. Some of the flagellin genes from different typing strains are identical: those from H40 and H8 are identical as are those from H15 and H16. Some others are nearly identical: H21 and H47 (1.5% difference), H12 and H1 (2.6% difference), H10 and H50 (0.3% difference), H38 and H55 (0.1% difference), H4, H44 and H17 are very similar, the pairwise difference ranging

from 0.33% to 0.87%.

In the cases where the flagellin gene from two type strains is near identical, we conclude that both genes code for flagellin of the same H specificity and that one or
 5 other strain has an additional locus which carries the functional gene, although the flagellin genes sequenced do not appear to be mutated.

As discussed above, genes encoding some H antigens have been shown to be located at loci other than *fliC*. H3,
 10 H36, H47, H53 have been shown to be at a locus called *flkA*, H44 and H55 at *fllA*, and H54 at *flmA* [Ratiner Y A (1998) "New flagellin-specifying genes in some *Escherichia coli* strains" J. Bacteriol. 180 979-984]. However, these strains may carry a *fliC* in addition to *flkA*, *fllA* or *flmA*
 15 [Ratiner Y A (1998) "New flagellin-specifying genes in some *Escherichia coli* strains" J. Bacteriol. 180 979-984].

The flagellin gene encoding H48 was previously sequenced from *E. coli* strain K-12 [Ku wajima G, Asaka J, Fujiwara T, Node K and Kondo E "Nucleotide sequence of the
 20 hag gene encoding flagellin of *Escherichia coli*" J Bacteriol. 168 (1986) 1479-1483]. We have sequenced the *fliC* gene from the H48 typing strain, and found that it is identical to that from K-12.

The H54 gene is known to be at *flmA* [Ratiner Y A
 25 (1998) "New flagellin-specifying genes in some *Escherichia coli* strains" J. Bacteriol. 180 979-984] and the finding of a non-functional presumptive *fliC* locus in the H54 strain shows that it is present but not expressed. However, we have not amplified and sequenced the
 30 functional *flmA* gene of this strain. The two bands from the H36 and H53 strains (both using primers based on *fliC* sequence) are thought to be from the *fliC* and *flkA* loci, but the bands were not purified and have not been sequenced. The *fliC* genes of H21 and H47 strains share

98.5% identity at DNA level and those of H38 and H55 strains share 99.9% identity at DNA level: as H47 and H55 map at *flkA* and *fllA* respectively we believe we have sequenced the *fliC* genes encoding the H21 and H38 antigenic specificities respectively and that these genes are present but not expressed in the H47 and H55 strains.

The genes encoding H36, H47 and H53 (at *flkA*), H44 and H55 (at *fllA*) and H54 (at *flmA*) are yet to be sequenced as those that were gave either 2 bands so the gene was not sequenced (H36 and H53), a nonfunctional gene (H54 and H35) or sequences obtained from the typing strains for these specificities are very similar to those of other strains (H47, H44 and H55) and we suspect that we have sequenced a non functional *fliC* gene in these cases. Also for other pairs with highly similar flagellin genes, as shown by our sequencing comparison, we do not know which H specificity we have sequenced leaving one still to be sequenced (H40 or H8, H15 or H16, H12 or H1, H10 or H50, and H4 or H17).

Using the 42 unique sequences and the sequences from the two non-functional flagellin genes (from H typing strains H35 and H54) (see Table 3) we have been able to determine antigen specific primers for each of the H antigen specificities and thereby show that it is practicable to detect *E.coli* strains carrying specific H antigens without false positives from strains of other H types. There is no reason to expect that the addition of 12 sequences to the 42 unique sequences obtained will affect the general conclusion, as unlike previous reports, our study covers flagellin sequences for a substantial majority of known *E. coli* H antigen specificities.

Our study of 11 H7 genes from strains of eight different O antigens shows limited variation and was such that the variation within genes for H antigens will not affect the ability to select antigen specific primers. O:H combinations in general define a strain and as some of the

strains thus defined were quite distant from each other in a study by Whittam [Whittam T S, wolfe M L, Wachsmuth I K, Orskov I and Wilson R A "Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea" Infect. Immun. 61 (1993) 1619-1629] the variation we observe is thought to represent that present in H7 genes. However, there is a low possibility that primers chosen without knowledge of the variation within genes of each H specificity could fail to give positive results with some isolates due to chance choice of primers which cover a base or bases which contribute to this low level variation. The variation within the H7 genes is in the normal range for variation within a gene in *E. coli* and if this possibility did occur it would be easy to use an alternate primer pair.

There are 54 known H antigens for *E. coli* and of these there are 12 H antigen specificities for which we do not as yet have sequence. It will be easy to determine these sequences and determine primer pairs specific for these H antigens by comparing these sequences with the 44 obtained sequences (see Table 3), and also modify the primers selected for any H antigen for which we already know the sequence in the unlikely event that there is a possibility of false positives with the primers selected.

The sequences for the remaining H antigens can be obtained in one of the following ways:

1. where we have two bands by PCR (H36 and H53 typing strains), we purify each and sequence, and also clone each into a strain mutated in its *fliC* gene and determine the H antigen expressed by use of specific sera. In this way a specific sequence can be related to an H antigen specificity. The other band which represents an H antigen gene for a different specificity is expected to include a mutant gene or a

gene similar to one of those already sequenced, but if not may represent a new specificity for which primer pairs could be selected. It may be difficult to obtain expression of flagellin genes when cloned from *E. coli* due to cloning together with regulatory sequences which prevent expression. This is easily avoided by cloning the major segment of the gene into a functioning *fliC* gene to replace the equivalent segment of that gene, using standard site directed mutagenesis to give suitable restriction sites within the cloned gene and incorporating those restriction sites into primers used to amplify the major segment of the gene to be studied to facilitate the cloning.

2. Where two or three strains have the same flagellin gene sequence, the genes are cloned as above and the H antigen specificity represented by this sequence is determined. This identifies the strain in which the gene is expressed and also those strains for which we have sequenced a gene which is not being expressed. We then clone the gene for the antigen expressed in these strains by making a bank of plasmid clones using chromosomal DNA and select for a clone which is expressing an H antigen different from the one represented by the known sequence. This can be done by taking advantage of the fact that the H antigen is on flagellin, the protein of the bacterial flagellum used for movement of the bacteria. In the presence of antibodies specific to that flagellum the bacteria cannot swim. For selection the clones are placed in a situation in which mobile cells can swim away from the others and be collected. There are many versions of these techniques and any could be used. One version is to place the bacteria on a nutrient agar plate with reduced agar content such that bacteria can swim away

from the site of inoculation. This is easily seen as growth on the plate and a sample of the bacteria which are motile can be recovered and cultivated. In this way bacteria carrying cloned H antigen genes can be
5 selected. If the medium in the plate has antibody added to it only bacteria which express an H antigen different to that recognised by the antiserum will be able to swim. Specifically if the antiserum used is specific for the H antigen expressed by the gene for
10 which we have sequence, only clones which express a different H antigen, such as those expressing the H antigen expressed by the H typing strains used to make the plasmid, will be selected. Once the clone is obtained, the H antigen gene can be sequenced.

15

Our work has shown that there are at least 8 cases where the H antigen typing strains carry two H antigen genes which appear to be complete and have the potential to function. However, while *E. coli* does not (in general)
20 have a capacity to express more than one flagellin gene, it is striking that there are several loci for flagellin genes [Ratiner Y A (1998) "New flagellin-specifying genes in some *Escherichia coli* strains" J. Bacteriol. 180 979-984]. Several of the pairs of H typing strains with identical
25 sequence do not include any of the H antigen types shown by Ratiner [Ratiner Y A (1998) "New flagellin-specifying genes in some *Escherichia coli* strains" J. Bacteriol. 180 979-984] to map other than at *fliC* although these predominate. This suggests that there are additional cases where an
30 expressed gene is not the only flagellin gene present. However the fact that flagellin gene sequences for many of the typing strains for H antigens found by Ratiner [Ratiner Y A (1998) "New flagellin-specifying genes in some *Escherichia coli* strains" J. Bacteriol. 180 979-984] to map
35 away from *fliC* are among those near identical to others,

indicates that the phenomenon is of limited extent. Nonetheless it remains possible even where only one gene has been obtained by PCR, that it is one of a pair of flagellin genes, the other not being amplified by the primers used, and further that it is the one not amplified which is expressing the H antigen of the strain. It will therefore be necessary to clone as described above each of the flagellin genes we have sequenced and confirm that it expresses the expected antigen to ensure that the invention give results corresponding to those of the traditional serotyping scheme. In the event that it does not, the gene for the type antigen can be cloned and sequenced by the means described above.

The 11 H7 *fliC* sequences fell into three groups, one comprising the genes from the O157:H7 and O55:H7 strains, which were identical, as expected given the proposed relationship between the clones. It has been shown that *E. coli* O157:H7 and O55:H7 clones are closely related [Whittam T S, wolfe M L, Wachsmuth I K, Orskov I and Wilson R A "Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea" Infect. Immun. 61 (1993) 1619-1629] thus it was expected that the H7 *fliC* genes from O157 and O55 would be identical. Among the H7 *fliC* sequences, we can identify primers specific to the H7 *fliC* gene for each of the three H7 groups. Two of these primers in combination with an H7 specific primer gave two primer pairs specific for the H7 gene of from the O157:H7 and O55:H7 clones.

6. Specific oligonucleotide primers for each of the 42 H types

Two oligonucleotide primers were chosen based on each of the 42 sequences. None of them had more than 85% identity with any other of 61 flagellin gene sequences. Thus,

these primers are specific for each H type. These primers are listed in Table 3.

The *fliC* gene of the H54 typing strain is a mutated gene. It has an insertion sequence (IS1222) inserted into a normal flagellin gene of H21. Thus, primers for H21 would amplify a fragment of different size in H54. We also provide 2 primers based on the insertion sequence (see H54 column in Table 3), the use of one of them in combination with one of the H21 primers will generate a PCR band only in H54, and this will also differentiate H54 from H21.

The *flic* gene of H35 type strain is also a mutated gene. It has an insertion sequence (IS1) inserted into a normal flagellin gene of H11. Thus, primers for H11 would amplify a fragment of different size in H35. We also provide 2 primers based on the insertion sequence (see H35 column in Table 3), the use of one of them in combination with one of the H11 primers will generate a PCR band only in H35, and this will also differentiate H35 from H11.

7. Testing of the H7 specific oligonucleotide primers

Primer pair #1806/#1809 (see Table 3) was used to carry out PCR on chromosomal DNA samples of all the 54 H type strains and the H7 strains listed in Table 1. PCR reactions were carried out under the following conditions: denaturing, 94°C/30'; annealing, 58°C/30'; extension, 72°C/1'; 30 cycles. PCR reaction was carried out in a volume of 50ul for each of the chromosomal sample. After the PCR reaction, 5ul PCR product from each sample was run on an agarose gel to check for amplified DNA.

Primer pairs #1806/#1809 produced a band of predicted size with all the 11 strains expressing H7, but gave no band with other H type strains. Thus,

these primers are H7 specific.

8. Testing of oligonucleotide primers specific to H7 of O157 and O55:

5 Based on a comparison of the *fliC* sequences of 11
different H7 strains, we have identified two
oligonucleotides [#1696 (5'-GGCCTGACTCAGGCGGCC) at
positions 178 to 195 in M527 and #1697 (5'-
10 GAGTTACCGGCCTGCTGA) positions 1700-1683 in M527] which
are unique to H7 of O157 and O55. Although not identical
to any parts of the *fliC* sequences of any other H7
strains, these two primers are identical or have high
level similarity to *fliC* genes of some other H types.
However a combination of one of these primers with one of
15 the H7 specific primers can give specificity for H7 of
O157 and O55 *E. coli*.

Primer pairs #1696/#1809 and #1697/#1806 were used
to carry out PCR on chromosomal DNA samples of all the H
type strains and the H7 strains listed in Table 1. PCR
20 reactions were carried out under the following
conditions: denaturing, 94°C/30'; annealing, 61°C/30'
(for #1696/#1809) or 60°C/30' (for #1697/#1806); extension,
72°C/1'; 30 cycles. PCR reaction was carried out in an
volume of 50µl for each of the chromosomal samples. After
25 the PCR reaction, 5µl PCR product from each sample was
run on an agarose gel to check for amplified DNA.

Both primer pairs produced a band of predicted size
with both of the H7:O157 strains (strains M1004 and M527,
see Table 1), and the H7:O55 strain (strain M1686, see
30 Table 1), gave no band with other strains. Thus, these
two pairs of primers are specific to H7 genes of O157 and
O55 *E. coli* strains.

O antigen

Materials and Methods-part 1

The experimental procedures for the isolation and characterisation of the *E. coli* O111 O antigen gene cluster (position 3,021-9,981) are according to Bastin D.A., et al. 1991 "Molecular cloning and expression in *Escherichia coli* K-12 of the *rfb* gene cluster determining the O antigen of an *E. coli* O111 strain". *Mol. Microbiol.* 5:9 2223-2231 and Bastin D.A. and Reeves, P.R. 1995 "Sequence and analysis of the O antigen gene(*rfb*)cluster of *Escherichia coli* O111". *Gene* 164: 17-23.

A. Bacterial strains and growth media

Bacteria were grown in Luria broth supplemented as required.

B. Cosmids and phage

Cosmids in the host strain x2819 were repackaged in vivo. Cells were grown in 250mL flasks containing 30mL of culture, with moderate shaking at 30°C to an optical density of 0.3 at 580 nm. The defective lambda prophage was induced by heating in a water bath at 45°C for 15min followed by an incubation at 37°C with vigorous shaking for 2hr. Cells were then lysed by the addition of 0.3mL chloroform and shaking for a further 10min. Cell debris were removed from 1mL of lysate by a 5min spin in a microcentrifuge, and the supernatant removed to a fresh microfuge tube. One drop of chloroform was added then shaken vigorously through the tube contents.

C. DNA preparation

Chromosomal DNA was prepared from bacteria grown overnight at 37°C in a volume of 30mL of Luria broth. After harvesting by centrifugation, cells were washed and resuspended in 10mL of 50mMTris-HCl pH 8.0. EDTA was added and the mixture incubated for 20min. Then lysozyme was added and incubation continued for a further 10min. Proteinase K, SDS, and ribonuclease were then added and the mixture incubated for up to 2hr for lysis to occur.

All incubations were at 37°C. The mixture was then heated to 65°C and extracted once with 8mL of phenol at the same temperature. The mixture was extracted once with 5mL of phenol/chloroform/iso-amyl alcohol at 4°C. Residual phenol was removed by two ether extractions. DNA was precipitated with 2 vols. of ethanol at 4°C, spooled and washed in 70% ethanol, resuspended in 1-2mL of TE and dialysed. Plasmid and cosmid DNA was prepared by a modification of the Birnboim and Doly method [Birnboim, H. C. and Doly, J. (1979) "A rapid alkaline extraction procedure for screening recombinant plasmid DNA" *Nucl. Acid Res.* 7:1513-1523]. The volume of culture was 10mL and the lysate was extracted with phenol/chloroform/iso-amyl alcohol before precipitation with isopropanol. Plasmid DNA to be used as vector was isolated on a continuous caesium chloride gradient following alkaline lysis of cells grown in 1L of culture.

D. Enzymes and buffers.

Restriction endonucleases and DNA T4 ligase were purchased from Boehringer Mannheim (Castle Hill, NSW, Australia) or Pharmacia LKB (Melbourne, VIC Australia). Restriction enzymes were used in the recommended commercial buffer.

E. Construction of a gene bank.

Individual aliquots of M92 chromosomal DNA (strain Stoke W, from Statens Serum Institut, 5 Artillerivej, 2300 Copenhagen S, Denmark) were partially digested with 0.2U *Sau3A*I for 1-15mins. Aliquots giving the greatest proportion of fragments in the size range of approximately 40-50kb were selected and ligated to vector pPR691 previously digested with *Bam*HI and *Pvu*II. Ligation mixtures were packaged *in vitro* with packaging extract. The host strain for transduction was x2819 and recombinants were selected with kanamycin.

F. Serological procedures.

Colonies were screened for the presence of the O111 antigen by immunoblotting. Colonies were grown overnight, up to 100 per plate then transferred to nitrocellulose
5 discs and lysed with 0.5N HCl. Tween 20 was added to TBS at 0.05% final concentration for blocking, incubating and washing steps. Primary antibody was *E. coli* O group 111 antiserum, diluted 1:800. The secondary antibody was goat anti-rabbit IgG labelled with horseradish peroxidase
10 diluted 1:5000. The staining substrate was 4-chloro-1-naphthol. Slide agglutination was performed according to the standard procedure.

G. Recombinant DNA methods.

Restriction mapping was based on a combination of
15 standard methods including single and double digests and sub-cloning. Deletion derivatives of entire cosmids were produced as follows: aliquots of 1.8mg of cosmid DNA were digested in a volume of 20ml with 0.25U of restriction enzyme for 5-80min. One half of each aliquot was used to
20 check the degree of digestion on an agarose gel. The sample which appeared to give a representative range of fragments was ligated at 4°C overnight and transformed by the CaCl₂ method into JM109. Selected plasmids were transformed into sf174 by the same method. P4657 was
25 transformed with pPR1244 by electroporation.

H. DNA hybridisation

Probe DNA was extracted from agarose gels by electroelution and was nick-translated using [α -³²P]-dCTP. Chromosomal or plasmid DNA was electrophoresed in 0.8%
30 agarose and transferred to a nitrocellulose membrane. The hybridisation and pre-hybridisation buffers contained either 30% or 50% formamide for low and high stringency probing respectively. Incubation temperatures were 42°C and 37°C for pre-hybridisation and hybridisation
35 respectively. Low stringency washing of filters consisted

of 3 x 20min washes in 2 x SSC and 0.1% SDS. High-stringency washing consisted of 3 x 5min washes in 2 x SSC and 0.1% SDS at room temperature, a 1hr wash in 1 x SSC and 0.1% SDS at 58°C and 15min wash in 0.1 x SSC and 0.1% SDS at 58°C.

I. Nucleotide sequencing of *E. coli* O111 O antigen gene cluster (position 3,021-9,981)

Nucleotide sequencing was performed using an ABI 373 automated sequencer (CA, USA). The region between map positions 3.30 and 7.90 was sequenced using uni-directional exonuclease III digestion of deletion families made in PT7T3190 from clones pPR1270 and pPR1272. Gaps were filled largely by cloning of selected fragments into M13mp18 or M13mp19. The region from map positions 7.90-10.2 was sequenced from restriction fragments in M13mp18 or M13mp19. Remaining gaps in both the regions were filled by priming from synthetic oligonucleotides complementary to determined positions along the sequence, using a single stranded DNA template in M13 or phagemid. The oligonucleotides were designed after analysing the adjacent sequence. All sequencing was performed by the chain termination method. Sequences were aligned using SAP [Staden, R., 1982 "Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing". *Nuc. Acid Res.* 10: 4731-4751; Staden, R., 1986 "The current status and portability of our sequence handling software". *Nuc. Acid Res.* 14: 217-231]. The program NIP [Staden, R. 1982 "An interactive graphics program for comparing and aligning nucleic acid and amino acid sequence". *Nuc. Acid Res.* 10: 2951-2961] was used to find open reading frames and translate them into proteins.

J. Isolation of clones carrying *E. coli* O111 O antigen gene cluster

The *E. coli* O antigen gene cluster was isolated according to the method of Bastin D.A., et al. [1991

"Molecular cloning and expression in *Escherichia coli* K-12 of the *rfb* gene cluster determining the O antigen of an *E. coli* O111 strain". *Mol. Microbiol.* 5(9), 2223-2231].

Cosmid gene banks of M92 chromosomal DNA were established in the *in vivo* packaging strain x2819. From the genomic bank, 3.3×10^3 colonies were screened with *E. coli* O111 antiserum using an immuno-blotting procedure: 5 colonies (pPR1054, pPR1055, pPR1056, pPR1058 and pPR1287) were positive. The cosmids from these strains were packaged *in vivo* into lambda particles and transduced into the *E. coli* deletion mutant Sf174 which lacks all O antigen genes. In this host strain, all plasmids gave positive agglutination with O111 antiserum. An *Eco* R1 restriction map of the 5 independent cosmids showed that they have a region of approximately 11.5 kb in common (Figure 1). Cosmid pPR1058 included sufficient flanking DNA to identify several chromosomal markers linked to O antigen gene cluster and was selected for analysis of the O antigen gene cluster region.

K. Restriction mapping of cosmid pPR1058

Cosmid pPR1058 was mapped in two stages. A preliminary map was constructed first, and then the region between map positions 0.00 and 23.10 was mapped in detail, since it was shown to be sufficient for O111 antigen expression. Restriction sites for both stages are shown in Figure 2. The region common to the five cosmid clones was between map positions 1.35 and 12.95 of pPR1058.

To locate the O antigen gene cluster within pPR1058, pPR1058 cosmid was probed with DNA probes covering O antigen gene cluster flanking regions from *S. enterica* LT2 and *E. coli* K-12. Capsular polysaccharide (*cps*) genes lie upstream of O antigen gene cluster while the gluconate dehydrogenase (*gnd*) gene and the histidine (*his*) operon are downstream, the latter being further from the O antigen gene cluster. The probes used were pPR472

(3.35kb), carrying the *gnd* gene of LT2, pPR685 (5.3kb) carrying two genes of the *cps* cluster, *cpsB* and *cpsG* of LT2, and K350 (16.5kb) carrying all of the *his* operon of K-12. Probes hybridised as follows: pPR472 hybridised to 1.55kb and 3.5 kb (including 2.7 kb of vector) fragments of *Pst*I and *Hind*III double digests of pPR1246 (a *Hind*III/*Eco*R1 subclone derived from pPR1058, Figure 2), which could be located at map positions 12.95-15.1; pPR685 hybridised to a 4.4 kb *Eco*R1 fragment of pPR1058 (including 1.3 kb of vector) located at map position 0.00-3.05; and K350 hybridised with a 32kb *Eco*R1 fragment of pPR1058 (including 4.0kb of vector), located at map position 17.30-45.90. Subclones containing the presumed *gnd* region complemented a *gnd*⁻*edd*⁻ strain GB23152. On gluconate bromothymol blue plates, pPR1244 and pPR1292 in this host strain gave the green colonies expected of a *gnd*⁺*edd*⁻ genotype. The *his*⁺ phenotype was restored by plasmid pPR1058 in the *his* deletion strain Sf174 on minimal medium plates, showing that the plasmid carries the entire *his* operon.

It is likely that the O antigen gene cluster region lies between *gnd* and *cps*, as in other *E. coli* and *S. enterica* strains, and hence between the approximate map positions 3.05 and 12.95. To confirm this, deletion derivatives of pPR1058 were made as follows: first, pPR1058 was partially digested with *Hind*III and self ligated. Transformants were selected for kanamycin resistance and screened for expression of O111 antigen. Two colonies gave a positive reaction. *Eco*R1 digestion showed that the two colonies hosted identical plasmids, one of which was designated pPR1230, with an insert which extended from map positions 0.00 to 23.10. Second pPR1058 was digested with *Sal*I and partially digested with *Xho*I and the compatible ends were re-ligated. Transformants were selected with kanamycin and screened for O111 antigen

expression. Plasmid DNA of 8 positively reacting clones was checked using *Eco*R1 and *Xho*I digestion and appeared to be identical. The cosmid of one was designated pPR1231. The insert of pPR1231 contained the DNA region between map positions 0.00 and 15.10. Third, pPR1231 was partially digested with *Xho*I, self-ligated, and transformants selected on spectinomycin/ streptomycin plates. Clones were screened for kanamycin sensitivity and of 10 selected, all had the DNA region from the *Xho*I site in the vector to the *Xho*I site at position 4.00 deleted. These clones did not express the O111 antigen, showing that the *Xho*I site at position 4.00 is within the O antigen gene cluster. One clone was selected and named pPR1288. Plasmids pPR1230, pPR1231, and pPR1288 are shown in Figure 2.

L. Analysis of the E. coli O111 O antigen gene cluster (position 3,021-9,981) nucleotide sequence data

Bastin and Reeves [1995 "Sequence and analysis of the O antigen gene(*rfb*)cluster of *Escherichia coli* O111". Gene 164: 17-23] partially characterised the *E. coli* O111 O antigen gene cluster by sequencing a fragment from map position 3,021-9,981. Figure 3 shows the gene organisation of position 3,021-9,981 of *E. coli* O111 O antigen gene cluster. *orf3* and *orf6* have high level amino acid identity with *wcaH* and *wcaG* (46.3% and 37.2% respectively), and are likely to be similar in function to sugar biosynthetic pathway genes in the *E. coli* K-12 colanic gene cluster. *orf4* and *orf5* show high levels of amino acid homology to *manC* and *manB* genes respectively. *orf7* shows high level homology with *rfbH* which is an abequose pathway gene. *orf8* encodes a protein with 12 transmembrane segments and has similarity in secondary structure to other *wzx* genes and is likely therefore to be the O antigen flippase gene.

Materials and Methods-part 2

A. Nucleotide sequencing of 1 to 3,020 and 9,982 to 14,516 of the *E. coli* O111 O antigen gene cluster

5 The sub clones which contained novel nucleotide sequences, pPR1231 (map position 0 and 1,510), pPR1237 (map position -300 to 2,744), pPR1239 (map position 2,744 to 4,168), pPR1245 (map position 9,736 to 12,007) and pPR1246 (map position 12,007 to 15,300) (Figure 2), were
10 characterised as follows: the distal ends of the inserts of pPR1237, pPR1239 and pPR1245 were sequenced using the M13 forward and reverse primers located in the vector. PCR walking was carried out to sequence further into each insert using primers based on the sequence data and the
15 primers were tagged with M13 forward or reverse primer sequences for sequencing. This PCR walking procedure was repeated until the entire insert was sequenced. pPR1246 was characterised from position 12,007 to 14,516. The DNA of these sub clones was sequenced in both directions. The
20 sequencing reactions were performed using the dideoxy termination method and thermocycling and reaction products were analysed using fluorescent dye and an ABI automated sequencer (CA, USA).

B. Analysis of the *E. coli* O111 O antigen gene cluster
25 (positions 1 to 3,020 and 9,982 to 14,516 of Figure 5) nucleotide sequence data

 The gene organisation of regions of *E. coli* O111 O antigen gene cluster which were not characterised by Bastin and Reeves [1995 "Sequence and analysis of the O
30 antigen gene(*rfb*)cluster of *Escherichia coli* O111." *Gene* 164: 17-23], (positions 1 to 3,020 and 9,982 to 14,516) is shown in Figure 3. There are two open reading frames in region 1. Four open reading frames are predicted in region 2. The position of each gene is listed in Table 9.

35 The deduced amino acid sequence of *orf1* (*wbdH*) shares

about 64% similarity with that of the *rfp* gene of *Shigella dysenteriae*. *Rfp* and *WbdH* have very similar hydrophobicity plots and both have a very convincing predicted transmembrane segment in a corresponding position. *rfp* is a galactosyl transferase involved in the synthesis of LPS core, thus *wbdH* is likely to be a galactosyl transferase gene. *orf2* has 85.7% identity at amino acid level to the *gmd* gene identified in the *E. coli* K-12 colanic acid gene cluster and is likely to be a *gmd* gene. *orf9* encodes a protein with 10 predicted transmembrane segments and a large cytoplasmic loop. This inner membrane topology is a characteristic feature of all known O antigen polymerases thus it is likely that *orf9* encodes an O antigen polymerase gene, *wzy*. *orf10* (*wbdL*) has a deduced amino acid sequence with low homology with *Lsi2* of *Neisseria gonorrhoeae*. *Lsi2* is responsible for adding GlcNAc to galactose in the synthesis of lipooligosaccharide. Thus it is likely that *wbdL* is either a colitose or glucose transferase gene. *orf11* (*wbdM*) shares high level nucleotide and amino acid similarity with *TrsE* of *Yersinia enterocolitica*. *TrsE* is a putative sugar transferase thus it is likely that *wbdM* encodes the colitose or glucose transferase.

In summary three putative transferase genes and an O antigen polymerase gene were identified at map position 1 to 3,020 and 9,982 to 14,516 of *E. coli* O111 O antigen gene cluster. A search of GenBank has shown that there are no genes with significant similarity at the nucleotide sequence level for two of the three putative transferase genes or the polymerase gene. Figure 5 provides the nucleotide sequence of the O111 antigen gene cluster.

Materials and Methods-part 3

A. PCR amplification of O157 antigen gene cluster from

an *E. coli* 0157:H7 strain (Strain C664-1992, from Statens Serum Institut, 5 Artillerivej, 2300, Copenhagen S, Denmark)

E. coli 0157 O antigen gene cluster was amplified by using long PCR [Cheng et al. 1994, "Effective amplification of long targets from cloned inserts and human and genomic DNA" P.N.A.S. USA 91: 5695-569] with one primer (primer #412: att ggt agc tgt aag cca agg gcg gta gcg t) based on the JumpStart sequence usually found in the promoter region of O antigen gene clusters [Hobbs, et al. 1994 "The JumpStart sequence: a 39 bp element common to several polysaccharide gene clusters" Mol. Microbiol. 12: 855-856], and another primer #482 (cac tgc cat acc gac gac gcc gat ctg ttg ctt gg) based on the *gnd* gene usually found downstream of the O antigen gene cluster. Long PCR was carried out using the Expand Long Template PCR System from Boehringer Mannheim (Castle Hill NSW Australia), and products, 14 kb in length, from several reactions were combined and purified using the Promega Wizard PCR preps DNA purification System (Madison WI USA). The PCR product was then extracted with phenol and twice with ether, precipitated with 70% ethanol, and resuspended in 40mL of water.

B. Construction of a random DNase I bank:

Two aliquots containing about 150ng of DNA each were subjected to DNase I digestion using the Novagen DNase I Shotgun Cleavage (Madison WI USA) with a modified protocol as described. Each aliquot was diluted into 45ml of 0.05M Tris -HCl (pH7.5), 0.05mg/mL BSA and 10mM MnCl₂. 5mL of 1:3000 or 1:4500 dilution of DNaseI (Novagen) (Madison WI USA) in the same buffer was added into each tube respectively and 10ml of stop buffer (100mM EDTA), 30% glycerol, 0.5% Orange G, 0.075% xylene and cyanol (Novagen) (Madison WI USA) was added after incubation at 15°C for 5 min. The DNA from the two DNaseI reaction

tubes were then combined and fractionated on a 0.8% LMT agarose gel, and the gel segment with DNA of about 1kb in size (about 1.5mL agarose) was excised. DNA was extracted from agarose using Promega Wizard PCR Preps DNA

5 Purification (Madison WI USA) and resuspended in 200 mL water, before being extracted with phenol and twice with ether, and precipitated. The DNA was then resuspended in 17.25 mL water and subjected to T4 DNA polymerase repair and single dA tailing using the Novagen Single dA Tailing
10 Kit (Madison WI USA). The reaction product (85ml containing about 8ng DNA) was then extracted with chloroform:isoamyl alcohol (24:1) once and ligated to 3×10^{-3} pmol pGEM-T (Promega) (Madison WI USA) in a total volume of 100mL. Ligation was carried out overnight at
15 4°C and the ligated DNA was precipitated and resuspended in 20mL water before being electroporated into *E. coli* strain JM109 and plated out on BCIG-IPTG plates to give a bank.

C. Sequencing

20 DNA templates from clones of the bank were prepared for sequencing using the 96-well format plasmid DNA miniprep kit from Advanced Genetic Technologies Corp (Gaithersburg MD USA) The inserts of these clones were sequenced from one or both ends using the standard M13
25 sequencing primer sites located in the pGEM-T vector. Sequencing was carried out on an ABI377 automated sequencer (CA USA) as described above, after carrying out the sequencing reaction on an ABI Catalyst (CA USA). Sequence gaps and areas of inadequate coverage were PCR
30 amplified directly from O157 chromosomal DNA using primers based on the already obtained sequencing data and sequenced using the standard M13 sequencing primer sites attached to the PCR primers.

35 D. Analysis of the *E. coli* O157 O antigen gene cluster nucleotide sequence data

Sequence data were processed and analysed using the Staden programs [Staden, R., 1982 "Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing." *Nuc. Acid Res.* 10: 4731-4751; Staden, R., 1986 "The current status and portability of our sequence handling software". *Nuc. Acid Res.* 14: 217-231; Staden, R. 1982 "An interactive graphics program for comparing and aligning nucleic acid and amino acid sequence". *Nuc. Acid Res.* 10: 2951-2961]. Figure 4 shows the structure of *E. coli* O157 O antigen gene cluster. Twelve open reading frames were predicted from the sequence data, and the nucleotide and amino acid sequences of all these genes were then used to search the GenBank database for indication of possible function and specificity of these genes. The position of each gene is listed in Table 9. The nucleotide sequence is presented in Figure 6.

orfs 10 and 11 showed high level identity to *manC* and *manB* and were named *manC* and *manB* respectively. *orf7* showed 89% identity (at amino acid level) to the *gmd* gene of the *E. coli* colanic acid capsule gene cluster (Stevenson G., K. et al. 1996 "Organisation of the *Escherichia coli* K-12 gene cluster responsible for production of the extracellular polysaccharide colanic acid". *J. Bacteriol.* 178:4885-4893) and was named *gmd*. *orf8* showed 79% and 69% identity (at amino acid level) respectively to *wcaG* of the *E. coli* colanic acid capsule gene cluster and to *wbcJ* (*orf14.8*) gene of the *Yersinia enterocolitica* O8 O antigen gene cluster (Zhang, L. et al. 1997 "Molecular and chemical characterization of the lipopolysaccharide O-antigen and its role in the virulence of *Y. enterocolitica* serotype O8". *Mol. Microbiol.* 23:63-76). Colanic acid and the *Yersinia* O8 O antigen both contain fucose as does the O157 O antigen. There are two

enzymatic steps required for GDP-L-fucose synthesis from GDP-4-keto-6-deoxy-D-mannose, the product of the *gmd* gene product. However, it has been shown recently (Tonetti, M et al. 1996 Synthesis of GDP-L-fucose by the human FX protein J. Biol. Chem. 271:27274-27279) that the human FX protein has "significant homology" with the *wcaG* gene (referred to as *Yefb* in that paper), and that the FX protein carries out both reactions to convert GDP-4-keto-6-deoxy-D-mannose to GDP-L-fucose. We believe that this makes a very strong case for *orf8* carrying out these two steps and propose to name the gene *fcl*. In support of the one enzyme carrying out both functions is the observation that there are no genes other than *manB*, *manC*, *gmd* and *fcl* with similar levels of similarity between the three bacterial gene clusters for fucose containing structures.

orf5 is very similar to *wbeE* (*rfbE*) of *Vibrio cholerae* O1, which is thought to be the perosamine synthetase, which converts GDP-4-keto-6-deoxy-D-mannose to GDP-perosamine (Stroeher, U.H et al. 1995 "A putative pathway for perosamine biosynthesis is the first function encoded within the *rfb* region of *Vibrio cholerae*" O1. Gene 166: 33-42). *V. cholerae* O1 and *E. coli* O157 O antigens contain perosamine and N-acetyl-perosamine respectively. The *V. cholerae* O1 *manA*, *manB*, *gmd* and *wbeE* genes are the only genes of the *V. cholerae* O1 gene cluster with significant similarity to genes of the *E. coli* O157 gene cluster and we believe that our observations both confirm the prediction made for the function of *wbe* of *V. cholerae*, and show that *orf5* of the O157 gene cluster encodes GDP-perosamine synthetase. *orf5* is therefore named *per*. *orf5* plus about 100bp of the upstream region (position 4022-5308) was previously sequenced by Bilge, S.S. et al. [1996 "Role of the *Escherichia coli* O157-H7 O side chain in adherence and analysis of an *rfb* locus". Infect.

Immun. 64:4795-4801].

orf12 shows high level similarity to the conserved region of about 50 amino acids of various members of an acetyltransferase family (Lin, W., et al. 1994 "Sequence analysis and molecular characterisation of genes required for the biosynthesis of type 1 capsular polysaccharide in *Staphylococcus aureus*". J. Bacteriol. 176: 7005-7016) and we believe it is the N-acetyltransferase to convert GDP-perosamine to GDP-perNAC. *orf12* has been named *wbdR*.

The genes *manB*, *manC*, *gmd*, *fcl*, *per* and *wbdR* account for all of the expected biosynthetic pathway genes of the 0157 gene cluster.

The remaining biosynthetic step(s) required are for synthesis of UDP-GalNac from UDP-Glc. It has been proposed (Zhang, L., et al. 1997 "Molecular and chemical characterisation of the lipopolysaccharide O-antigen and its role in the virulence of *Yersinia enterocolitica* serotype O8". Mol. Microbiol. 23:63-76) that in *Yersinia enterocolitica* UDP-GalNac is synthesised from UDP-GlcNac by a homologue of galactose epimerase (*GalE*), for which there is a *galE* like gene in the *Yersinia enterocolitica* O8 gene cluster. In the case of 0157 there is no *galE* homologue in the gene cluster and it is not clear how UDP-GalNac is synthesised. It is possible that the galactose epimerase encoded by the *galE* gene in the *gal* operon, can carry out conversion of UDP-GlcNac to UDP-GalNac in addition to conversion of UDP-Glc to UDP-Gal. There do not appear to be any gene(s) responsible for UDP-GalNac synthesis in the 0157 gene cluster.

orf4 shows similarity to many *wzx* genes and is named *wzx* and *orf2* which shows similarity of secondary structure in the predicted protein to other *wzy* genes and is for that reason named *wzy*.

The *orf1*, *orf3* and *orf6* gene products all have

characteristics of transferases, and have been named *wbdN*,
wbdO and *wbdP* respectively. The O157 O antigen has 4
sugars and 4 transferases are expected. The first
transferase to act would put a sugar phosphate onto
undecaprenol phosphate. The two transferases known to
perform this function, WbaP (RfbP) and WecA (Rfe) transfer
galactose phosphate and N-acetyl-glucosamine phosphate
respectively to undecaprenol phosphate. Neither of these
sugars is present in the O157 structure.

Further, none of the presumptive transferases in the
O157 gene cluster has the transmembrane segments found in
WecA and WbaP which transfer a sugar phosphate to
undecaprenol phosphate and expected for any protein which
transferred a sugar to undecaprenol phosphate which is
embedded within the membrane.

The WecA gene which transfers GlcNAc-P to
undecaprenol phosphate is located in the Enterobacterial
Common Antigen (ECA) gene cluster and it functions in ECA
synthesis in most and perhaps all *E. coli* strains, and
also in O antigen synthesis for those strains which have
GlcNAc as the first sugar in the O unit.

It appears that WecA acts as the transferase for
addition of GalNAc-1-P to undecaprenol phosphate for the
Yersinia enterocolitica O8 O antigen [Zhang et al.1997
"Molecular and chemical characterisation of the
lipopolysaccharide O antigen and its role in the virulence
of *Yersinia enterocolitica* serotype O8" Mol. Microbiol.
23: 63-76.] and perhaps does so here as the O157 structure
includes GalNAc. WecA has also been reported to add
Glucose-1-P phosphate to undecaprenol phosphate in *E. coli*
O8 and O9 strains, and an alternative possibility for
transfer of the first sugar to undecaprenol phosphate is
WecA mediated transfer of glucose, as there is a glucose
residue in the O157 O antigen. In either case the
requisite number of transferase genes are present if

GalNAc or Glc is transferred by *WecA* and the side chain Glc is transferred by a transferase outside of the O antigen gene cluster.

5 *orf9* shows high level similarity (44% identity at amino acid level, same length) with *wcaH* gene of the *E. coli* colanic acid capsule gene cluster. The function of this gene is unknown, and we give *orf9* the name *wbdQ*.

10 The DNA between *manB* and *wdbR* has strong sequence similarity to one of the H-repeat units of *E. coli* K12. Both of the inverted repeat sequences flanking this region are still recognisable, each with two of the 11 bases being changed. The H-repeat associated protein encoding gene located within this region has a 267 base deletion and mutations in various positions. It seems that the H-repeat unit has been associated with this gene cluster for
15 a long period of time since it translocated to the gene cluster, perhaps playing a role in assembly of the gene cluster as has been proposed in other cases.

20 Materials and Methods - part 4

25 To test our hypothesis that O antigen genes for transferases and the *wzx*, *wzy* genes were more specific than pathway genes for diagnostic PCR, we first carried out PCR using primers for all the *E. coli* 016 O antigen genes (Table 7). The PCR was then carried out using PCR
30 primers for *E. coli* 0111 transferase, *wzx* and *wzy* genes (Table 8, 8A). PCR was also carried out using PCR primers for the *E. coli* 0157 transferase, *wzx* and *wzy* genes (Table 9, 9A).

35 Chromosomal DNA from the 166 serotypes of *E. coli* available from Statens Serum Institut, 5 Artillerivej, 2300 Copenhagen Denmark was isolated using the Promega Genomic (Madison WI USA) isolation kit. Note that 164 of the serogroups are described by Ewing W. H.: Edwards and
40 Ewings "Identification of the Enterobacteriaceae" Elsevier,

Amsterdam 1986 and that they are numbered 1-171 with numbers 31, 47, 67, 72, 93, 94 and 122 no longer valid. Of the two serogroup 19 strains we used 19ab strain F8188-41. Lior H. 1994 ["Classification of *Escherichia coli* In *Escherichia coli* in domestic animals and humans pp 31-72. Edited by C.L. Gyles CAB international] adds two more numbered 172 and 173 to give the 166 serogroups used. Pools containing 5 to 8 samples of DNA per pool were made. Pool numbers 1 to 19 (Table 4) were used in the *E. coli* 0111 and 0157 assay. Pool numbers 20 to 28 were also used in the 0111 assay, and pool numbers 22 to 24 contained *E. coli* 0111 DNA and were used as positive controls (Table 5). Pool numbers 29 to 42 were also used in the 0157 assay, and pool numbers 31 to 36 contained *E. coli* 0157 DNA, and were used as positive controls (Table 6). Pool numbers 2 to 20, 30, 43 and 44 were used in the *E. coli* 016 assay (Tables 4 to 6). Pool number 44 contained DNA of *E. coli* K-12 strains C600 and WG1 and was used as a positive control as between them they have all of the *E. coli* K-12 016 O antigen genes.

PCR reactions were carried out under the following conditions: denaturing 94°C/30"; annealing, temperature varies (refer to Tables)/30"; extension, 72°C/1'; 30 cycles. PCR reaction was carried out in a volume of 25mL for each pool. After the PCR reaction, 10mL PCR product from each pool was run on an agarose gel to check for amplified DNA.

Each *E. coli* chromosomal DNA sample was checked by gel electrophoresis for the presence of chromosomal DNA and by PCR amplification of the *E. coli mdh* gene using oligonucleotides based on *E. coli* K-12 [Boyd et al. (1994) "Molecular genetic basis of allelic polymorphism in malate dehydrogenase (*mdh*) in natural populations of *Escherichia coli* and *Salmonella enterica*" Proc. Nat. Acad. Sci. USA. 91:1280-1284.] Chromosomal DNA samples from other

bacteria were only checked by gel electrophoresis of chromosomal DNA.

A. Primers based on *E. coli* O16 O antigen gene cluster sequence.

The O antigen gene cluster of *E. coli* O16 was the only typical *E. coli* O antigen gene cluster that had been fully sequenced prior to that of O111, and we chose it for testing our hypothesis. One pair of primers for each gene was tested against pools 2 to 20, 30 and 43 of *E. coli* chromosomal DNA. The primers, annealing temperatures and functional information for each gene are listed in Table 8.

For the five pathway genes, there were 17/21, 13/21, 0/21, 0/21, 0/21 positive pools for *rmlB*, *rmlD*, *rmlA*, *rmlC* and *glf* respectively (Table 7). For the *wzx*, *wzy* and three transferase genes there were no positives amongst the 21 pools of *E. coli* chromosomal DNA tested (Table 7). In each case the #44 pool gave a positive result.

B. Primers based on the *E. coli* O111 O antigen gene cluster sequence.

One to four pairs of primers for each of the transferase, *wzx* and *wzy* genes of O111 were tested against the pools 1 to 21 of *E. coli* chromosomal DNA (Table 8). For *wbdH*, four pairs of primers, which bind to various regions of this gene, were tested and found to be specific for O111 as there was no amplified DNA of the correct size in any of those 21 pools of *E. coli* chromosomal DNA tested. Three pairs of primers for *wbdM* were tested, and they are all specific although primers #985/#986 produced a band of the wrong size from one pool. Three pairs of primers for *wzx* were tested and they all were specific. Two pairs of primers were tested for *wzy*, both are

specific although #980/#983 gave a band of the wrong size in all pools. One pair of primers for *wbdL* was tested and found unspecific and therefore no further test was carried out. Thus, *wzx*, *wzy* and two of the three transferase genes are highly specific to 0111. Bands of the wrong size found in amplified DNA are assumed to be due to chance hybridisation of genes widely present in *E. coli*. The primers, annealing temperatures and positions for each gene are in Table 8.

The 0111 assay was also performed using pools including DNA from O antigen expressing *Yersinia pseudotuberculosis*, *Shigella boydii* and *Salmonella enterica* strains (Table 8A). None of the oligonucleotides derived from *wbdH*, *wzx*, *wzy* or *wbdM* gave amplified DNA of the correct size with these pools. Notably, pool number 25 includes *S. enterica* Adelaide which has the same O antigen as *E. coli* 0111: this pool did not give a positive PCR result for any primers tested indicating that these genes are highly specific for *E. coli* 0111.

Each of the 12 pairs binding to *wbdH*, *wzx*, *wzy* and *wbdM* produces a band of predicted size with the pools containing 0111 DNA (pools number 22 to 24). As pools 22 to 24 included DNA from all strains present in pool 21 plus 0111 strain DNA (Table 5), we conclude that the 12 pairs of primers all give a positive PCR test with each of three unrelated 0111 strains but not with any other strains tested. Thus these genes are highly specific for *E. coli* 0111.

C. Primers based on the *E. coli* 0157 O antigen gene cluster sequence.

Two or three primer pairs for each of the transferase, *wzx* and *wzy* genes of 0157 were tested against *E. coli* chromosomal DNA of pools 1 to 19, 29 and 30 (Table

9). For *wbdN*, three pairs of primers, which bind to various regions of this gene, were tested and found to be specific for O157 as there was no amplified DNA in any of those 21 pools of *E. coli* chromosomal DNA tested. Three
5 pairs of primers for *wbdO* were tested, and they are all specific although primers # 1211/#1212 produced two or three bands of the wrong size from all pools. Three pairs of primers were tested for *wbdP* and they all were specific. Two pairs of primers were tested for *wbdR* and
10 they were all specific. For *wzy*, three pairs of primers were tested and all were specific although primer pair #1203/#1204 produced one or three bands of the wrong size in each pool. For *wzx*, two pairs of primers were tested and both were specific although primer pair #1217/#1218
15 produced 2 bands of wrong size in 2 pools, and 1 band of wrong size in 7 pools. Bands of the wrong size found in amplified DNA are assumed to be due to chance hybridisation of genes widely present in *E. coli*. The primers, annealing temperatures and function information
20 for each gene are in Table 9.

The O157 assay was also performed using pools 37 to 42, including DNA from O antigen expressing *Yersinia pseudotuberculosis*, *Shigella boydii*, *Yersinia enterocolitica* 09, *Brucella abortus* and *Salmonella*
25 *enterica* strains (Table 9A). None of the oligonucleotides derived from *wbdN*, *wzy*, *wbdO*, *wzx*, *wbdP* or *wbdR* reacted specifically with these pools, except that primer pair #1203/#1204 produced two bands with *Y. enterocolitica* 09 and one of the bands is of the same size with that from
30 the positive control. Primer pair #1203/#1204 binds to *wzy*. The predicted secondary structures of Wzy proteins are generally similar, although there is very low similarity at amino acid or DNA level among the sequenced *wzy* genes. Thus, it is possible that *Y. enterocolitica* 09

has a *wzy* gene closely related to that of *E. coli* 0157. It is also possible that this band is due to chance hybridization of another gene, as the other two *wzy* primer pairs (#1205/#1206 and #1207/#1208) did not produce any
5 band with *Y. enterocolitica* 09. Notably, pool number 37 includes *S. enterica* Landau which has the same O antigen as *E. coli* 0157, and pool 38 and 39 contain DNA of *B. abortus* and *Y. enterocolitica* 09 which cross react serologically with *E. coli* 0157. This result indicates
10 that these genes are highly 0157 specific, although one primer pair may have cross reacted with *Y. enterocolitica* 09.

Each of the 16 pairs binding to *wbdN*, *wzx*, *wzy*, *wbdO*, *wbdP* and *wbdR* produces a band of predicted size with the
15 pools containing 0157 DNA (pools number 31 to 36). As pool 29 included DNA from all strains present in pools 31 to 36 other than 0157 strain DNA (Table 6), we conclude that the 16 pairs of primers all give a positive PCR test with each of the five unrelated 0157 strains.

20 Thus PCR using primers based on genes *wbdN*, *wzy*, *wbdO*, *wzx*, *wbdP* and *wbdR* is highly specific for *E. coli* 0157, giving positive results with each of six unrelated 0157 strains while only one primer pair gave a band of the expected size with one of three strains with O antigens
25 known to cross-react serologically with *E. coli* 0157.

TABLE 1

H7 strains used in this work in addition to the typing strains

Name used in this study	Serotype	Original name	Source*
M527	O157:H7	C664-1992	a
M917	O18ac:H7	A57	IMVS
M918	O18ac:H7	A62	IMVS
M973	O2:H7	A1107	CDC
M1004	O157:H7	EH7	b
M1179	O18ac:H7	D-M3291/54	IMVS
M1200	O7:H7	A64	c
M1211	O19ab:H7	F8188-41	IMVS
M1328	O53:H7	14097	IMVS

*

a. Statens Serum Institut, Copenhagen, Denmark.

b. Dr. R. Brown of Royal Children's Hospital, Melbourne, Australia.

c. Max-Planck Institut fur molekulare Genetik, Berlin, Germany.

d. Dr. P. Tarr of Children's Hospital and Medical Center, University of Washington, USA

IMVS Institute of Medical and Veterinary Science, Adelaide, Australia.

CDC Centers for Disease Control and Prevention, Atlanta, USA

TABLE 2
Oligonucleotides used to PCR amplify *fliC* genes
from different H typing strains

H Type	Annealing Temperature (°C)	Primers Used
1	55	#1575/#1576
2	55	#1285/#1286
3	55	#1285/#1286
4	50	#1431/#1432
5	60	#1285/#1286
6	55	#1575/#1576
7	55	#1575/#1576
8	55	#1431/#1432
9	60	#1575/#1576
10	55	#1575/#1576
11	55	#1285/#1286
12	60	#1575/#1576
14	60	#1575/#1576
15	60	#1575/#1576
16	60	#1575/#1576
17	60	#1417/#1418
18	60	#1575/#1576
19	60	#1575/#1576
20	60	#1575/#1576
21	55	#1285/#1286
23	60	#1575/#1576
24	60	#1285/#1286
25	60	#1417/#1418
26	60	#1575/#1576
27	50	#1431/#1432
28	60	#1575/#1576
29	60	#1285/#1286
30	60	#1575/#1576
31	60	#1575/#1576
32	60	#1575/#1576
33	60	#1285/#1286
34	55	#1575/#1576
35	50	#1431/#1432
37	60	#1285/#1286
38	60	#1285/#1286
39	55	#1285/#1286
40	55	#1285/#1286
41	60	#1575/#1576
42	60	#1285/#1286
43	60	#1575/#1576
44	60	#1285/#1286
45	60	#1575/#1576
46	60	#1575/#1576
47	55	#1285/#1286
48	60	#1575/#1576
49	60	#1575/#1576
50	60	#1285/#1286
51	60	#1575/#1576
52	60	#1575/#1576
54	50	#1431/#1432
55	60	#1285/#1286
56	60	#1285/#1286

TABLE 3 Specific H type oligonucleotide primers

H type	Flagellin gene sequence used for primer choice	Positions of primer 1	Positions of primer 2	Other flagellin gene(s) highly similar to this	Loci other than flhC encoding the H antigen
1	No (obtained from GenBank)			12	
2	yes	568-587	1039-1056		
3	yes	649-666	925-942		flhA
4	yes	466-483	628-648	44,17	
5	yes	697-714	877-897		
6	yes	565-585	799-816		
7	yes	553-570 (primer #1806)	1483-1500 (primer #1809)		
8	yes	562-579	1045-1062	40	
9	yes	616-633	838-855		
10	yes	559-579	697-717	50	
11	yes	586-606*	790-810*		
12	yes	745-765	1024-1041	1	
14	yes	586-606	793-813		
15	yes	640-660	817-834	16	
16	No			15	
17	No			4 and 44	
18	yes	589-606	802-819		
19	yes	607-624	838-855		
20	yes	574-591	760-780		
21	yes	676-693**	862-879**	47	
23	yes	637-654	1336-1353		
24	yes	496-516	772-792		
25	yes	529-549	703-723		
26	yes	553-570	772-789		
27	yes	685-702	799-819		
28	yes	592-609	778-798		
29	yes	538-555	757-774		
30	yes	814-831	943-962		
31	yes	571-588	790-807		
32	yes	814-831	1057-1074		
33	yes	553-570	718-735		
34	yes	568-585	796-816		
35	no (non-functional gene)	769-789*	1045-1065*		
36	No (PCR generated two bands)				flhA
37	yes	520-537	715-735		
38	yes	553-573	709-729	55	
39	yes	556-573	718-735		
40	No			8	
41	yes	598-615	784-801		
42	yes	547-567	715-735		
43	yes	580-597	844-861		
44	No			4 and 17	flhA
45	yes	640-657	943-963		
46	yes	565-582	781-801		
47	No			21	flhA
48	yes	568-585	835-852		
49	yes	589-609	754-771		
50	No			10	
51	yes	565-582	1042-1059		
52	yes	598-615	829-846		
53	No (PCR generated two bands)				flhA
54	No (non-functional gene)	988-1008**	1344-1364**		flmA
55	No			38	flhA
56	yes	697-714	877-897		

* See text for choice of primers for flhC gene of H11

** See text for choice of primers for flhC gene of H21

TABLE 4

Pool No.	Strains of which chromosomal DNA included in the pool	Source*
1	<i>E. coli</i> type strains for O serotypes 1, 2, 3, 4, 10, 16, 18 and 39	IMVS ^a
2	<i>E. coli</i> type strains for O serotypes 40, 41, 48, 49, 71, 73, 88 and 100	IMVS
3	<i>E. coli</i> type strains for O serotypes 102, 109, 119, 120, 121, 125, 126 and 137	IMVS
4	<i>E. coli</i> type strains for O serotypes 138, 139, 149, 7, 5, 6, 11 and 12	IMVS
5	<i>E. coli</i> type strains for O serotypes 13, 14, 15, 17, 19ab, 20, 21 and 22	IMVS
6	<i>E. coli</i> type strains for O serotypes 23, 24, 25, 26, 27, 28, 29 and 30	IMVS
7	<i>E. coli</i> type strains for O serotypes 32, 33, 34, 35, 36, 37, 38 and 42	IMVS
8	<i>E. coli</i> type strains for O serotypes 43, 44, 45, 46, 50, 51, 52 and 53	IMVS
9	<i>E. coli</i> type strains for O serotypes 54, 55, 56, 57, 58, 59, 60 and 61	IMVS
10	<i>E. coli</i> type strains for O serotypes 62, 63, 64, 65, 66, 68, 69 and 70	IMVS
11	<i>E. coli</i> type strains for O serotypes 74, 75, 76, 77, 78, 79, 80 and 81	IMVS
12	<i>E. coli</i> type strains for O serotypes 82, 83, 84, 85, 86, 87, 89 and 90	IMVS
13	<i>E. coli</i> type strains for O serotypes 91, 92, 95, 96, 97, 98, 99 and 101	IMVS
14	<i>E. coli</i> type strains for O serotypes 103, 104, 105, 106, 107, 108 and 110	IMVS
15	<i>E. coli</i> type strains for O serotypes 112, 162, 113, 114, 115, 116, 117 and 118	IMVS
16	<i>E. coli</i> type strains for O serotypes 123, 165, 166, 167, 168, 169, 170 and 171	See b
17	<i>E. coli</i> type strains for O serotypes 172, 173, 127, 128, 129, 130, 131 and 132	See c
18	<i>E. coli</i> type strains for O serotypes 133, 134, 135, 136, 140, 141, 142 and 143	IMVS
19	<i>E. coli</i> type strains for O serotypes 144, 145, 146, 147, 148, 150, 151 and 152	IMVS

*

- a. Institute of Medical and Veterinary Science, Adelaide, Australia
- b. 123 from IMVS; the rest from Statens Serum Institut, Copenhagen, Denmark
- c. 172 and 173 from Statens Serum Institut, Copenhagen, Denmark, the rest from IMVS

TABLE 5

Pool No.	Strains of which chromosomal DNA included in the pool	Source*
20	<i>E. coli</i> type strains for O serotypes 153, 154, 155, 156, 157, 158, 159 and 160	IMVS
21	<i>E. coli</i> type strains for O serotypes 161, 163, 164, 8, 9 and 124	IMVS
22	As pool #21, plus <i>E. coli</i> 0111 type strain Stoke W.	IMVS
23	As pool #21, plus <i>E. coli</i> 0111:H2 strain C1250-1991	See d
24	As pool #21, plus <i>E. coli</i> 0111:H12 strain C156-1989	See e
25	As pool #21, plus <i>S. enterica</i> serovar Adelaide	See f
26	<i>Y. pseudotuberculosis</i> strains of O groups IA, IIA, IIB, IIC, III, IVA, IVB, VA, VB, VI and VII	See g
27	<i>S. boydii</i> strains of serogroups 1, 3, 4, 5, 6, 8, 9, 10, 11, 12, 14 and 15	See h
28	<i>S. enterica</i> strains of serovars (each representing a different O group) Typhi, Montevideo, Ferruch, Jangwari, Raus, Hvitvingfoss, Waycross, Dan, Dugbe, Basel, 65:i:e,n,z,15 and 52:d:e,n,x,z15	IMVS

*

- d. C1250-1991 from Statens Serum Institut, Copenhagen, Denmark
- e. C156-1989 from Statens Serum Institut, Copenhagen, Denmark
- f. *S. enterica* serovar Adelaide from IMVS
- g. Dr S Aleksic of Institute of Hygiene, Germany
- h. Dr J Lefebvre of Bacterial Identification Section, Laboratoire de Santé Publique du Québec, Canada

TABLE 6

Pool No.	Strains of which chromosomal DNA included in the pool	Source*
29	<i>E. coli</i> type strains for O serotypes 153, 154, 155, 156, 158, 159 and 160	IMVS
30	<i>E. coli</i> type strains for O serotypes 161, 163, 164, 8, 9, 111 and 124	IMVS
31	As pool #29, plus <i>E. coli</i> O157 type strain A2 (O157:H19)	IMVS
32	As pool #29, plus <i>E. coli</i> O157:H16 strain C475-89	See d
33	As pool #29, plus <i>E. coli</i> O157:H45 strain C727-89	See d
34	As pool #29, plus <i>E. coli</i> O157:H2 strain C252-94	See d
35	As pool #29, plus <i>E. coli</i> O157:H39 strain C258-94	See d
36	As pool #29, plus <i>E. coli</i> O157:H26	See e
37	As pool #29, plus <i>S. enterica</i> serovar Landau	See f
38	As pool #29, plus <i>Brucella abortus</i>	See g See h
39	As pool #29, plus <i>Y. enterocolitica</i> O9	
40	<i>Y. pseudotuberculosis</i> strains of O groups IA, IIA, IIB, IIC, III, IVA, IVB, VA, VB, VI and VII	See i
41	<i>S. boydii</i> strains of serogroups 1, 3, 4, 5, 6, 8, 9, 10, 11, 12, 14 and 15	See j
42	<i>S. enterica</i> strains of serovars (each representing a different O group) Typhi, Montevideo, Ferruch, Jangwani, Raus, Hvittingfoss, Waycross, Dan, Dugbe, Basel, 65:i:e,n,z15 and 52:d:e,n,x,z15	IMVS
43	<i>E. coli</i> type strains for O serotypes 1,2,3,4,10,18 and 29	IMVS
44	As pool #43, plus <i>E. coli</i> K-12 strains C600 and WG1	IVMS See k

*

- d. O157 strains from Statens Serum Institut, Copenhagen, Denmark
- e. O157:H26 from Dr R Brown of Royal Children's Hospital, Melbourne, Victoria
- f. *S. enterica* serovar Landau from Dr M Poppoff of Institut Pasteur, Paris, France
- g. *B. Abortus* from the culture collection of The University of Sydney, Sydney, Australia
- h. *Y. enterocolitica* O9 from Dr. K. Bettelheim of Victorian Infectious Diseases Reference Laboratory Victoria, Australia.
- i. Dr S Aleksic of Institute of Hygiene, Germany
- J. Dr J Lefebvre of Bacterial Identification Section, Laboratoire de Santé Publique du Québec, Canada
- k. Strains C600 and WG1 from Dr. B.J. Backmann of Department of Biology, Yale University, USA.

TABLE 7 PCR assay result using primers based on the *E. coli* serotype O16 (strain K-12) O antigen gene cluster sequence

Gene	Function	Base positions of the gene	Forward primer (base positions)	Reverse primer (base positions)	Length of the PCR fragment	Number of pools (out of 21) giving band of correct size	Annealing temperature of the PCR
<i>rmlB</i> *	TDP-rhamnose pathway	90-1175	#1064(91-109)	#1065(1175-1157)	1085bp	17	60°C
<i>rmlD</i> *	TDP-rhamnose pathway	1175-2074	#1066(1175-1193)	#1067 (2075-2058)	901bp	13	60°C
<i>rmlA</i> *	TDP-rhamnose pathway	2132-3013	#1068(2131-2148)	#1069(3013-2995)	883bp	0	60°C
<i>rmlC</i> *	TDP-rhamnose pathway	3013-3570	#1070(3012-3029)	#1071(3570-3551)	559bp	0	60°C
<i>glt</i> *	Galactofuranose pathway	4822-5925	#1074(4822-4840)	#1075(5925-5908)	1104bp	0	55°C
<i>wzx</i> *	Flippase	3567-4814	#1072(3567-3586)	#1073(4814-4797)	1248bp	0	55°C
<i>wzy</i> *	O polymerase	5925-7091	#1076(5925-5944)	#1077(7091-7074)	1167bp	0	60°C
<i>wbbI</i> *	Galactofuranosyl transferase	7094-8086	#1078 (7094-7111)	#1079(8086-8069)	993bp	0	50°C
<i>wbbJ</i> *	Acetyltransferase	8067-8654	#1080(8067-8084)	#1081(8654-8632)	588bp	0	60°C
<i>wbbK</i> **	Glucosyl transferase	5770-6888	#1082(5770-5787)	#1083(6888-6871)	1119bp	0	55°C
<i>wbbL</i> ***	Rhamnosyltransferase	679-1437	#1084(679-697)	#1085(1473-1456)	795bp	0****	55°C

*, **, *** Base positions based on GenBank entry U09876, U03041 and L19537 respectively
**** 19 pools giving a band of wrong size

TABLE 8 PCR assay data using 0111 primers

Gene	Base positions of the gene according to SEQ ID NO: 1	Forward primer (base positions)	Reverse primer (base positions)	Length of the PCR fragment	Number of pools (out of 21) giving band of correct size	Annealing temperature of the PCR
<i>wbdH</i>	739-1932	#866 (739-757)	#867(1941-1924)	1203bp	0	60°C
		#976(925-942)	#978(1731-1714)	807bp	0	60°C
		#976(925-942)	#979(1347-1330)	423bp	0	60°C
		#977(1165-1182)	#978(1731-1714)	567bp	0	60°C
<i>wzx</i>	8646-9911	#969(8646-8663)	#970(9908-9891)	1263bp	0	50°C
		#1060(8906-8923)	#1062(9468-9451)	563bp	0	60°C
		#1061(9150-9167)	#1063 (9754-9737)	605bp	0	50°C
<i>wzy</i>	9901-10953	#900(9976-9996)	#901(10827-10807)	852bp	0	60°C
		#980(10113-10130)	#983(10484-10467)	372bp	0*	61°C
<i>wbdL</i>	10931-11824	#870(10931-10949)	#871(11824-11796)	894bp	7	60°C
<i>wbdM</i>	11821-12945	#868(11821-11844)	#869(12945-12974)	1125bp	0	60°C
		#984(12042-12059)	#987(12447-12430)	406bp	0	60°C
		#985(12258-12275)	#986(12698-12681)	441bp	0**	65°C

* Giving a band of wrong size in all pools
 ** One pool giving a band of wrong size

TABLE 8A PCR specificity test data using 0111 primers

Gene	Base positions of the gene according to SEQ ID NO: 1	Forward primer (base positions)	Reverse primer (base positions)	Length of the PCR fragment	Number of pools (pools no. 25-28) giving band of correct size	Annealing temperature of the PCR
<i>wbdH</i>	739-1932	#866 (739-757)	#867 (1941-1924)	1203bp	0*	60°C
		#976 (925-942)	#978 (1731-1714)	807bp	0	60°C
		#976 (925-942)	#979 (1347-1330)	423bp	0	60°C
		#977 (1165-1182)	#978 (1731-1714)	567bp	0	60°C
<i>wzx</i>	8646-9911	#969 (8646-8663)	#970 (9908-9891)	1263bp	0	55°C
		#1060 (8906-8923)	#1062 (9468-9451)	563bp	0	60°C
		#1061 (9150-9167)	#1063 (9754-9737)	605bp	0*	50°C
		#900 (9976-9996)	#901 (10827-10807)	852bp	0	60°C
<i>wzy</i>	9901-10953	#980 (10113-10130)	#983 (10484-10467)	372bp	0**	60°C
		#870 (10931-10949)	#871 (11824-11796)	894bp	0	60°C
<i>wbdL</i>	10931-11824	#868 (11821-11844)	#869 (12945-12924)	1125bp	0	60°C
<i>wbdM</i>	11821-12945	#984 (12042-12059)	#987 (12447-12430)	406bp	0	60°C
		#985 (12258-12275)	#986 (12698-12681)	441bp	0*	65°C

* 1 pool giving a band of wrong size
 ** 2 pools giving 3 bands of wrong sizes, 1 pool giving 2 bands of wrong sizes

TABLE 9 PCR results using primers based on the *E. coli* O157 sequence

Gene	Function	Base position of the gene according to SEQ ID NO: 2	Forward primer (base positions)	Reverse primer (base positions)	Length of the PCR fragment	Number of pools (out of 21) giving band of correct size	Annealing temperature of the PCR
<i>wbdN</i>	Sugar transferase	79-861	#1197(79-96)	#1198 (861-844)	783	0	55°C
			#1199(184-201)	#1200(531-514)	348	0	55°C
			#1201(310-327)	#1202(768-751)	459	0	55°C
<i>wzy</i>	O antigen	858-2042	#1203(858-875)	#1204(2042-2025)	1185	0*	50°C
			#1205(1053-1070)	#1206(1619-1602)	567	0	63°C
			#1207(1278-1295)	#1208(1913-1896)	636	0	60°C
<i>wbdO</i>	Sugar transferase	2011-2757	#1209(2011-2028)	#1210(2757-2740)	747	0	50°C
			#1211(2110-2127)	#1212(2493-2476)	384	0**	62°C
			#1213(2305-2322)	#1214(2682-2665)	378	0	60°C
<i>wzx</i>	O antigen flippase	2744-4135	#1215(2744-2761)	#1216(4135-4118)	1392	0	50°C
			#1217(2942-2959)	#1218(3628-3611)	687	0***	63°C
<i>wbdP</i>	Sugar transferase	5257-6471	#1221(5257-5274)	#1222(6471-6454)	1215	0	55°C
			#1223(5440-5457)	#1224(5973-5956)	534	0	55°C
			#1225(5707-5724)	#1226(6231-6214)	525	0	55°C
<i>wbdR</i>	N-acetyl	13156-13821	#1229(13261-13278)	#1230(13629-13612)	369	0	55°C
			#1231(13384-13401)	#1232(13731-13714)	348	0	60°C

* 3 bands of wrong size in one pool, 1 band of wrong size in all other pools

** 3 bands of wrong sizes in 9 pools, 2 bands of wrong size in all other pools

*** 2 bands of wrong sizes in 2 pools, 1 band of wrong size in 7 pools

TABLE 9A PCR results using primers based on the *E. coli* O157 sequence

Gene	Function	Base position of the gene according to SEQ ID NO: 2	Forward primer (base positions)	Reverse primer (base positions)	Length of the PCR fragment	Number of pools (pools no. 37-42) giving band of correct size	Annealing temperature of the PCR
<i>wbdN</i>	Sugar transferase	79-861	#1197(79-96)	#1198 (861-844)	783	0*	55°C
			#1199(184-201)	#1200(531-514)	348	0*	55°C
			#1201(310-327)	#1202(768-751)	459	0	61°C
<i>wzy</i>	O antigen	858-2042	#1203(858-875)	#1204(2042-2025)	1185	1**	50°C
			#1205(1053-1070)	#1206(1619-1602)	567	0***	60°C
			#1207(1278-1295)	#1208(1913-1896)	636	0	60°C
<i>wbdO</i>	Sugar transferase	2011-2757	#1209(2011-2028)	#1210(2757-2740)	747	0	50°C
			#1211(2110-2127)	#1212(2493-2476)	384	0****	61°C
			#1213(2305-2322)	#1214(2682-2665)	378	0	60°C
<i>wzx</i>	O antigen flippase	2744-4135	#1215(2744-2761)	#1216(4135-4118)	1392	0	50°C
			#1217(2942-2959)	#1218(3628-3611)	687	0	63°C
<i>wbdP</i>	Sugar transferase	5257-6471	#1221(5257-5274)	#1222(6471-6454)	1215	0	55°C
			#1223(5440-5457)	#1224(5973-5956)	534	0*	60°C
			#1225(5707-5724)	#1226(6231-6214)	525	0	55°C
<i>wbdR</i>	N-acetyl transferase	13156-13821	#1229(13261-13278)	#1230(13629-13610)	369	0	50°C
			#1231(13384-13401)	#1232(13731-13710)	348	0	60°C

-39-

- * 1 band of wrong size in one pool
- ** pool #39 giving two bands, one band of correct size, the other band of wrong size in another pool.
- *** 2 bands of wrong sizes in one pool
- **** 3 bands of wrong sizes in 2 pools, 2 bands of wrong sizes in 2 other pools

CLAIMS:

1. A nucleic acid molecule encoding all or part of an *E. coli* flagellin protein.
2. A method of detecting the presence of *E. coli* of a particular H serotype in a sample, the method comprising the step of specifically hybridising at least one nucleic acid molecule derived from a flagellin gene, wherein the at least one nucleic acid molecule is specific for a particular flagellin gene associated with the H serotype, to any *E. coli* in the sample which contain the gene, and detecting any specifically hybridised nucleic acid molecules, wherein the presence of specifically hybridised nucleic acid molecules identifies the presence of the H serotype in the sample.
3. A method of detecting the presence of *E. coli* of a particular H serotype in a sample, the method comprising the step of specifically hybridising at least one pair of nucleic acid molecules to any *E. coli* in the sample which contains the flagellin gene for the particular H serotype, wherein at least one of the nucleic acid molecules is specific for the particular flagellin gene associated with the H serotype, and detecting any specifically hybridised nucleic acid molecules, wherein the presence of specifically hybridised nucleic acid molecules identifies the presence of the H serotype in the sample.
4. A method for detecting the presence of a particular O serotype and H serotype of *E. coli* in a sample, the method comprising the following steps:
 - (a) specifically hybridising at least one nucleic acid molecule, derived from and specific for a gene

encoding a transferase or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit, the gene being involved in the synthesis of a particular *E. coli* O antigen, to any *E. coli* in the sample which contain the gene;

(b) specifically hybridising at least one nucleic acid molecule derived from and specific for a particular flagellin gene associated with that H serotype, to any *E. coli* in the sample which contain the gene; and

(c) detecting any specifically hybridised nucleic acid molecules.

5. A method for detecting the presence of a particular O serotype and H serotype of *E. coli* in a sample, the method comprising the following steps:

(a) specifically hybridising at least one pair of nucleic acid molecules, derived from and specific for a gene encoding a transferase or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit, the gene being involved in the synthesis of a particular *E. coli* O antigen, to any *E. coli* in the sample which contain the gene;

(b) specifically hybridising at least one pair of nucleic acid molecules derived from and specific for a particular flagellin gene associated with that H serotype, to any *E. coli* in the sample which contain the gene; and

(c) detecting any specifically hybridised nucleic acid molecules.

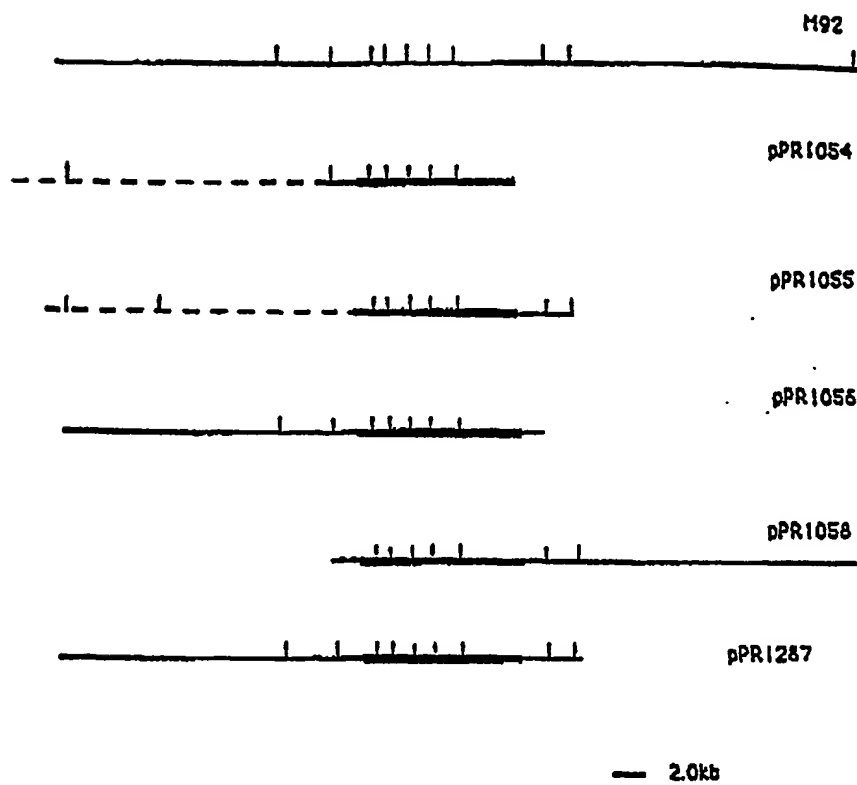


Figure 1

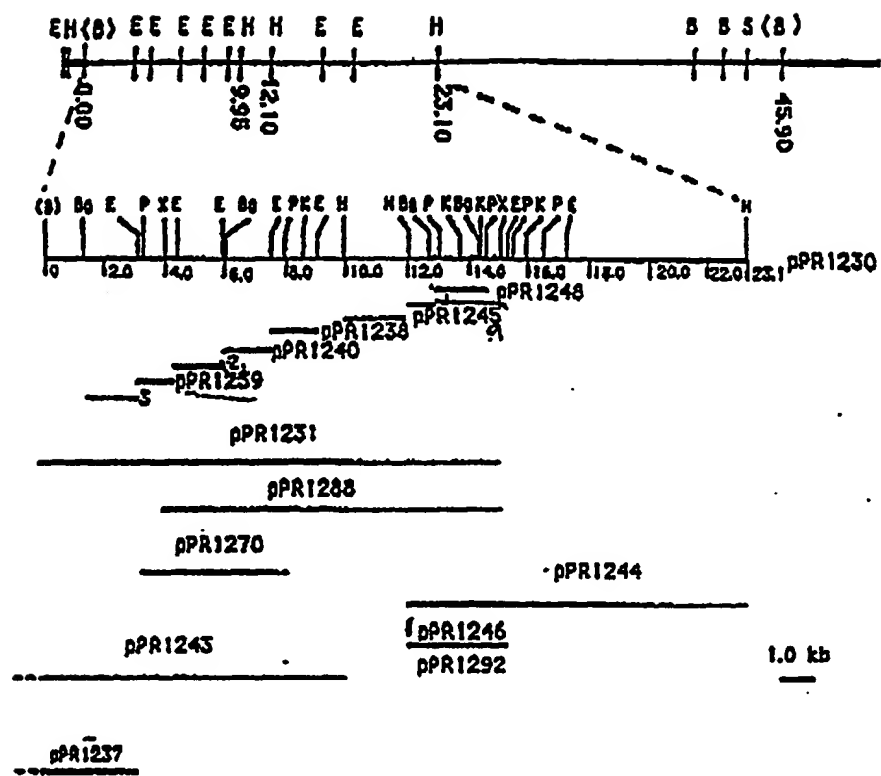


Figure 2

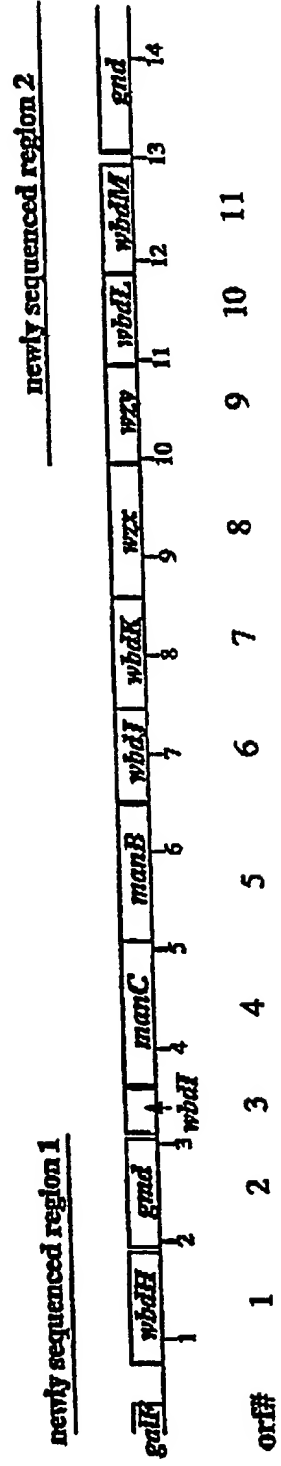


Figure 3

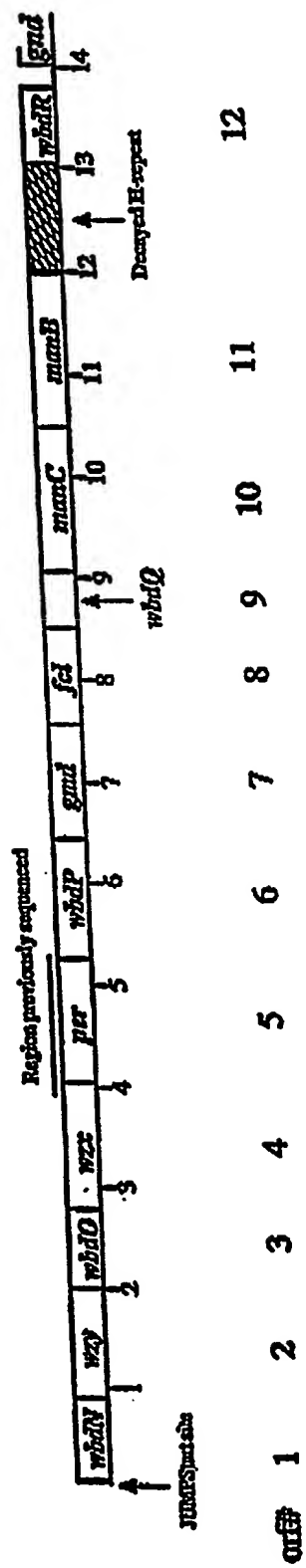


Figure 4

GATCTGATGGCCGTAGGGCGCTACGTGCTTTCTGCTGATATCTGGGCTGAGTTGGAAAAA	60
ACTGCTCCAGGTGCCGTGGGGACGTATTCAACTGACTGATGCTATTGCAGAGTTGGCTAAA	120
AAACAGTCTGTTGATGCCATGCTGATGACCGCGACAGCTACGACTGCGGTAAGAAGATG	180
GGCTATATGCAGGCATTTCGTTAAGTATGGGCTGCGCAACCTTAAAGAAGGGGCGAAGTTC	240
CGTAAGAGCATCAAGAAGCTACTGAGTGAGTAGAGATTTACACGTCTTTGTGACGATAAG	300
CCAGAAAAAATAGCGGCAGTTAACATCCAGGCTTCTATGCTTTAAGCAATGGAATGTTAC	360
TGCCGTTTTTTTATGAAAAATGACCAATAATAACAAGTTAACCTACCAAGTTTAATCTGCT	420
TTTTGTGGATTTTTTCTTGTTCCTGGTCGCATTTGGTAAGACAAATTAGCGTGAGTTTTA	480
GAGAGTTTTGCGGGATCTCGCGGAAGTGTCTCACATCTTTGGCATTTAGTTAGTGCCTGG	540
TAGCTGTTAAGCCAGGGGCGGTAGCTTGCCTAATTAATTTTAAAGTATACATTTATTCT	600
TGCCGCTTATAGCAAATAAAGTCAATCGGATTAAACTTCTTTTCCATTAGGTAAAAGAGT	660
GTTTGTAGTCGCTCAGGGAAATTGGTTTTGGTAGTAGTACTTTTCAAATTATCCATTTTC	720
<p style="text-align: center;">Start of orf1</p>	
<p style="text-align: center;">M L L C C I H I N V Y Y L L</p> CGATTTAGATGGCAGTTGATGTTACTATGCTGCATACATCAATGTATATTATTACTT	780
<p style="text-align: center;">L E C D M K K I V I I G N V A S M M L R</p> TTAGAATGTGATATGAAAAAATAGTGATCATAGGCAATGTAGCGTCAATGATGTTAAGG	840
<p style="text-align: center;">F R K E L I M N L V R Q G D N V Y C L A</p> TTCAGGAAGAATTAATCATGAATTTAGTGAGGCAAGGTGATAATGTATATTGTCTAGCA	900
<p style="text-align: center;">N D F S T E D L K V L S S W G V K G V K</p> AATGATTTTTCCACTGAAGATCTTAAAGTACTTTCGTCATGGGGCGTTAAGGGGGTTAAA	960
<p style="text-align: center;">F S L N S K G I N P F K D I I A V Y E L</p> TTCTCTCTTAACTCAAAGGGTATTATCCTTTTAAAGGATATAATTGCTGTTTATGAACTA	1020
<p style="text-align: center;">K K I L K D I S P D I V F S Y F V K P V</p> AAAAAATTCCTTAAGGATATTTCCCCAGATATTGTATTTTCATATTTTGTAAAGCCAGTA	1080
<p style="text-align: center;">I F G T I A S K L S K V P R I V G M I E</p> ATATTTGGAATATTGCTTCAAAGTTGTCAAAGTGCCAAGGATTGTTGGAATGATTGAA	1140
<p style="text-align: center;">G L G N A F T Y Y K G K Q T T K T K M I</p> GGTCTAGGTAATGCCTTCACTTATTATAAGGGAAAGCAGACCACAAAACTAAAATGATA	1200
<p style="text-align: center;">K W I Q I L L Y K L A L P M L D D L I L</p> AAGTGATACAAATTCCTTTTATATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTA	1260
<p style="text-align: center;">L N H D D K K D L I D Q Y N I K A K V T</p> TTAAATCATGATGATAAAAAAGATTTAATCGATCAGTATAATATTAAGCTAAGGTAACA	1320
<p style="text-align: center;">V L G G I G L D L N E F S Y K E P P K E</p> GTGTTAGGTGGGATTGGATTGGATCTTAATGAGTTTTCATATAAAGAGCCACCGAAAGAG	1380
<p style="text-align: center;">K I T F I F I A R L L R E K G I F E F I</p> AAAATTACCTTTATTTTTATAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTATT	1440
<p style="text-align: center;">E A A K F V K T T Y P S S E F V I L G G</p> GAAGCCGCAAAGTTCGTTAAGACAACCTATCCAAGTCTGAATTTGTAATTTTAGGAGGT	1500

Figure 5/1

F E S N N P F S L Q K N E I E S L R K E TTTGAGAGTAATAATCCTTTCTCATTACAAAAAATGAAATTGAATCGCTAAGAAAAGAA	1560
H D L I Y P G H V E N V Q D W L E K S S CATGATCTTATTTATCCTGGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTTC	1620
V F V L P T S Y R E G V P R V I Q E A M GTTTTGTGTTTACCTACATCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATG	1680
A I G R P V I T T N V P G C R D I I N D GCTATTGGTAGACCTGTAATAACAATAATGTACCTGGGTGTAGGGATATAATAAATGAT	1740
G V N G F L I P P F E I N L L A E K M K GGGGTCAATGGCTTTTTGATACCTCCATTGAAATTAATTTACTGGCAGAAAAAATGAAA	1800
Y F I E N K D K V L E M G L A G R K F A TATTTTATTGAGAATAAAGATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTGCA	1860
E K N F D A F E K N N R L A S I I K S N GAAAAAACTTTGATGCTTTTGAAAAAATAATAGACTAGCATCAATAATAAATCAAAT	1920
 End of orf1	
N D F *	
AATGATTTTTGACTTGAGCAGAAATTATTTATATTTCAATCTGAAAAATAAAGGCTGTTA	1980
 Start of orf2	
M N K V A L I T G I T G Q D G S Y L A TTATGAATAAAGTGGCATTAACTACTGGTATCACTGGGCAAGATGGCTCCTATTGGCAG	2040
E L L L E K G Y E V H G I K R R A S S F AATTATTGTTAGAAAAAGGTTATGAAGTTCATGGTATTAAACGCCGTGCATCTTCATTTA	2100
N T E R V D H I Y Q D S H L A N P K L F ATACTGAGCGAGTGGATCACATCTATCAGGATTCACATTAGCTAATCCTAAACTTTTTTC	2160
L H Y G D L T D T S N L T R I L K E V Q TACACTATGGCGATTGACAGATACTTCCAATCTGACCCGTATTTTAAAAGAAGTTCAAC	2220
P D E V Y N L G A M S H V A V S F E S P CAGATGAAGTTTACAATTTGGGGCGATGAGCCATGTAGCGGTATCATTTGAGTCACCAG	2280
E Y T A D V D A I G T L R L L E A I R I AATACACTGCTGATGTTGATGCGATAGGAACATTGCGTCTTCTTGAAGCTATCAGGATAT	2340
L G L E K K T K F Y Q A S T S E L Y G L TGGGGCTGAAAAAAGACAAATTTTATCAGGCTTCAACTTCAGAGCTTTATGGTTTGG	2400
V Q E I P Q K E T T P F Y P R S P Y A V TTCAAGAAATTCACAAAAAGAGACTACGCCATTTTATCCAGTTTCGCCTTATGCTGTTG	2460
A K L Y A Y W I T V N Y R E S Y G M F A CAAAATATATGCCTATTGGATCACTGTAAATTATCGTGAGTCTTATGGTATGTTTGCCT	2520
C N G I L F N H E S P R R G E T F V T R GCAATGGTATTCTCTTTAACCACGAATCACCTCGCCGTGGCGAGACCTTTGTTACTCGTA	2580
K I T R G I A N I A Q G L D K C L Y L G AAATAACACGCGGGATAGCAAATATTGCTCAAGGTCTTGATAAATGCTTATACTTGGGAA	2640
N M D S L R D W G H A K D Y V K M Q W M ATATGGATTCTCTGCGTGATTGGGGACATGCTAAGGATTATGTCAAAATGCAATGGATGA	2700

Figure 5/2

M L Q Q E T P E D F V I A T G I Q Y S V
TGCTGCAGCAAGAACTCCAGAAGATTTTGTAAATTGCTACAGGAATTCATATTTCTGTCC 2760
R E F V T M A A E Q V G I E L A F E G E
GTGAGTTTGTCACAATGGCGGCAGAGCAAGTAGGCATAGAGTTAGCATTGGAAGGTGAGG 2820
G V N E K G V V V S V N G T D A K A V N
GAGTAAATGAAAAAGGTGTTGTTGTTTCGGTCAATGGCACTGATGCTAAAGCTGTAAACC 2880
P G D V I I S V D P R Y F R P A E V E T
CGGGCGATGTAATTATATCTGTAGATCCAAGGTATTTTAGGCCTGCAGAAGTTGAAACCT 2940
L L G D P T N A H K K L G W S P E I T L
TGCTTGGCGATCCTACTAATGCGCATAAAAAATTAGGATGGAGCCCTGAAATTACATTGC 3000
R E M V K E M V S S D L A I A K K N V L
GTGAAATGGTAAAAGAAATG6TTTTCAGCGATTAGCAATAGCGAATAAGAAAGT6TTTTC 3060

End of orf2
L K A N N I A T N I P Q E *
TGAAGCTAATAACATTGCGCACTAATATTCGCAAGAAATAAAAAGATAATACATTAAT 3120

Start of orf3
M F
AATTAAAAATGGTCTAGATTATATTAGTACCATTATTTTTCGGTCACTAATGTTTA 3180
I T S D K F R E I I K L V P L V S I D L
TTACATCAGATAAATTTAGAGAAATTATCAAGTTAGTTCCATTAGTATCAATTGATCTGC 3240
L I E N E N G E Y L F G L R N N R P A K
TAATTGAAAACGAGAATGGTCAATATTTATTTGCTCTTAGGAATAATCGACCGGCGAAAA 3300
N Y F F V P G G R I R K N E S I K N A F
ATTAATTTTTTTTGTTCAGGTGGTAGGATTCGCAAAATGAAATCTATTAAATAATGCTTTTA 3360
K R I S S M E L G K E Y G I S G S V F N
AAAGAAATATCATCTATGGAATTAGGTAAGAGTATGCTATTTTCAGGAAGTGTTTTTTAATG 3420
G V W E H F Y D D G F F S E G E A T H Y
GTGTATGGGAACATTTCTATGATGATGTTTTCCTGTAAGGCGAGGCAACACATATA 3480
I V L C Y T L K V L K S E L N L P D D Q
TAGTGTCTTTGTTACACACTGAAAGTTCTTAAAGTGAATTGAATCTCCACAGATGATCAAC 3540
H R E Y L W L T K H Q I N A K Q D V H N
ATCGTGAATACCTTTGCGTAACAAACCAATAATGCTAAACAAAGATCTTCATAACT 3600

End of orf3 Start of orf4
Y S K N Y F L * M
ATTCAAAAATTAATTTTTTGTAAATTTTATTAATAATTAATATGCGAGAGAAATTGTATGT 3660
S Q C L Y P V I I A G G T G S R L W P L
CTCAATGTCTTTTACCTCTAATTTATGCGCGAGGAACCGGAAGCGCTCTATGCGCGTTGT 3720
S R V L Y P K Q F L N L V G D S T M L Q
CTCGAGTATTATACCTAAACAAATTTTAAATTTAGTTGGGGATTCTACAAATGTTGCAAA 3780
T T I T R L D G I E C E N P I V I C N E
CAACAATTACCGCTTTGCGATGCGATCGAATGCGAAATTECAATTGTTATCTGCAATGAAG 3840
D H R F I V A E Q L R Q I G K L T K N I
ATCAACGATTTATTTGAGCAGAGCAATTACGACAGATTCGTAAGCTAACCAAGCAATATTA 3900
I L E P K G R N T A P A I A L A A F I A
TACTTGAGCGGAAAGGCGGTAATCTGCACTGCGCATAGCTTTAGCTGCTTTTATGCTGTC 3960

Figure 5/3

Q K N N P N D D P L L L V L A A D H S I 4020
AGAAGATAATCCTAATGACGACCCCTTTATTATTAGTACTTGGGGCAGACCACTCTATAA
N N E K A F R E S I I K A M P Y A T S G 4080
ATAATGAAAAAGCATTTCGAGAGTCAATAATAAAGCTATGCEGTATGCAACTTCTGCGA
K L V T F G I I P D T A N T G Y G Y I K 4140
AGTTAGTAACATTTCGAATTATTTCGGACACGGCAATACTGGTTATGGATATATTAAGA
R S S S A D P N K E F P A Y N V A E F V 4200
GAAGTTCTTCAGCTGATCCTAATAAGAATTCCGAGCATATAATGTTGCGGAGTTTGTAG
E K P D V K T A Q E Y I S S G N Y Y W N 4260
AAAACCCAGATGTTAAACAGGACAGGAATATATTTTCGAGTGGGAATTATTACTGGAATA
S G M F L F R A S K Y L D E L R K F R P 4320
GGCGAATGTTTTTATTTCGGCGCCAGTAAATATCTTGTATGAACTACGGAAATTTAGACCAG
D I Y H S C E C A T A T A N I D M D F V 4380
ATATTTATCATAGCTGTGAATGTGCAACCGCTACAGCAATATAGATATGGACTTTCTCC
R I N E A E F I N C P E E S I D Y A V M 4440
GAATTAACGAGGCTGAGTTTATTAAATGCTCTGAGAGCTCTATGCAATTATGCTGTGATGG
E K T K D A V V L P I D I G W N D V G S 4500
AAAAACAAAAGACGCTGTAGTTCTTCGGATAGATATTGGCTGGAATGACGTGGGTTCTT
W S S L W D I S Q K D C H G N V C H G D 4560
GGTCATCACTTTGGGATATAAGCCAAAGGATTGCCATGGTAAATGTGTGCCATGGGGATG
V L N H D G E N S F I Y S E S S L V A T 4620
TGCTCAATCATGATGGAGAAAATAGTTTTTATTACTCTGAGTCAACTCTGGTTCCGACAG
V G V S N L V I V Q T K D A V L V A D R 4680
TCGGAGTAAGTAATTTAGTAATTTGTCCAAACCAAGGATGCTGTACTGCTTGGGACCGTG
D K V Q N V K N I V D D L K K R K R A E 4740
ATAAGTCCAAAATGTTAAAAACATAGTTGACGATCTAAAAAGAGAAAACGTGCTGAAT
Y Y M H R A V F R P W G K F D A I D Q G 4800
ACTACATGCATCCTGCACTTTTTCGCGCTTGGGGTAAATTCGATGCAATAGACCAAGGCG
D R Y R V K K I I V K P G E G L D L R M 4860
ATAGATATAGAGTAAAAAAATAATAGTTAAACCAGGAGAGGGTTAGATTAAAGGATGC
H H H R A E H W I V V S G T A K V S L G 4920
ATCATCATAGGGCAGAGCATTGCAATTGTTGTATCCGGTACTGCTAAGTTTCACTAGGTA
S E V K L L V S N E S I Y I P Q G A K Y 4980
GTGAAGTTAACTATTAGTTTCTAATGAGTCTATATATATATCCCTCAGGGAGCAAAATATA
S L E N P G V I P L H L I E V S S G D Y 5040
GTCTTCAGAAATCCAGCGCTAATACCTTTGCATCTAATTCAGTAAGTTCTGCTGATTACC
L E S D D I V R F T D R Y N S K Q F L K 5100
TTGAAATCAGATGATATACTGCGTTTTTACTGACAGATATAACAGTAAACAAATTCCTAAGC

End of orf4 Start of orf5

M N K I T C F K A Y D I R G R L
R D *
GAGATGATAAATATGAATAAATAAATTTGCTTCAAGCATATGATATACTGGGGCTCT 5160

Figure 5/4

G A E L N D E I A Y R I G R A Y G E F F TGGTGGCTGAATTGAATGATGAATAGCATATAGAATTGGTGGCGCTTATGGTGAAGTTTTT	5220
K P Q T V V V G G D A R L T S E S L K K TAAACCTCAAACTGTAGTTGTGGGAGGAGATGCTGCGCTTAAACAAGTGAGAGTTTAAAGAA	5280
S L S N G L C D A G V N V L D L G M C G ATCACTCTCAAAATGGGCTATGTGATGCGAGGCGTAAATGTCCTTAGATCTTGGAAATGTGTGG	5340
T E E I Y F S T W Y L G I D G G I E V T TACTGAAGAGATATATTTTTCACCTTGGTATTTAGGAATTGATGGTGGAAATCGAGGTAAC	5400
A S H N P I D Y N G M K L V T K G A R P TGCAGGCCATAATCCAAATTGATTATTAATGGAAATGAATTTAGTAACCAAAAGGTGCTCGACC	5460
I S S D T G L K D I Q Q L V E S N N F E AATCAGCAGTCACACAGGTCTCAAGATATACAACAATTAGTAGAGAGTAATAATTTTGA	5520
E L N L E K K G N I T K Y S T R D A Y I AGAGCTCAACCTAGAAAAAAGGGAATATTACCAAAATTTCCACCCGAGATGCGCTACAT	5580
N H L M G Y A N L Q K I K K I K I V V N AAATCATTTGATGGGCTATGCTAATCTGCAAAAAATAAAAAAATCAAAATAGTTGTGAA	5640
S G N G A A G P V I D A I E E C F L R N TTCTGGGAATGGTGGCAGCTGGTCTCTGTTATTGATGCTATTGAGGAATGCTTTTTTACGGAA	5700
N I P I Q F V K I N N T P D G N F P H G CAATATTCCGATTACGTTTGTAAAAATAAATAATACACCCGATGCTAATTTTCCACATGG	5760
I P N P L L P E C R E D T S S A V I R H TATCCCTAATCCATTACTACCTGAGTGCAGAGAAGATAACCAGCAGTGGCGTTATAAGACA	5820
S A D F G I A F D G D F D R C F F F D E TACTGCTGATTTTGGTATTGCAATTGATGGTGAATTTTGATAGGTGTTTTTTCTTTGATGA	5880
N G Q F I E G Y Y I V G L L A E V F L G AAATGGACAAATTTATTGAAGGATACTACATTGTTGCTTTATTAGCGGAAGTTTTTTTAGG	5940
K Y P N A K I I H D P R L I W N T I D I GAAATATCCAAACGCAAAATCATTCATGATCTCTGCGCTTATATGGAATTAATTTGATAT	6000
V E S H G G I P I M T K T G H A Y I K Q CCTAGAAAGTCATGGTGGTATACCTATAATGACTAAACCGGTCATGCTTACATTAAACA	6060
R M R E E D A V Y G G E M S A H H Y F K AAGAAATGCGTGAAGAGGATGCGGTATATGGCGGCGAAATGAGTGGCGCATCATTATTTTAA	6120
D F A Y C D S G M I P W I L I C E L L S AGATTTTGCATACTGCGATAGTGGAAATGATTCTTGGATTTTAATTTGTGAACTTTTCAG	6180
L T N K K L G E L V C G C I N D W P A S TCTGACAAATAAAAAATTAGGTGAACGTGTTTGTGGTTGTATAAACGACTGGCGGCGAAG	6240
G E I N C T L D N P Q N E I D K L F N R TGGAGAAATAAACTGTACACTAGACAATCEGCAAAATGAATAGATAAATTATTTAATCG	6300
Y K D S A L A V D Y T D G L T M E F S D TTACAAAGATAGTGGCTTAGCTGTGATTACACTGATGGATTAACTATGGAGTTCTCTGA	6360
W R F N V R C S N T E P V V R L N V E S TTGGCGTTTAAATGTTAGATGCTCAATAACAGAACTGTAGTACGATGGAATGTAGAATE	6420
R N N A I L M Q E K T E E I L N F I S K TAGGAATAATGCTATTCTTATGCGAGAAAAACAGAGAAATCTGAATTTTATATCAAA	6480

Figure 5/5

End of orf5	Start of orf6	
*	M K V L L T G	
ATAAATTTGCACCTGAGTTCATAATGGGAACAAGAAATATATGAAAGTACTTCTGACTGG		6540
S T G M V G K N I L E H D S A S K Y N I		
ETCAACTGGCATGGTTGGTAAGAAATATATTAGAGCATGATACTGCCAAGTAAATATAATAT		6600
L T P T S S D L N L L D K N E I E K F M		
ACTTACTCCAACCAGCTCTGATTGAAATTTATTAGATAAAATGAAATAGAAAATTCAT		6660
L I N M P D C I I H A A G L V G G I H A		
GCTTATCAACATGCCAGACTGTATTATACATGCCAGCGGATTAGTTGGAGGCATTTCATGC		6720
N I S R P F D F L E K N L Q M G L N L V		
AAATATAAGCAGGCGCGTTTCTGTTTCTGGAATAAAATTTGCAGATGGGTTTAAATTTAGT		6780
S V A K K L G I K K V L N L G S S C M Y		
TTCCGTCGCAAAAATACTAGGTATCAAGAAAGTCTTAACTTGGGTAGTTTCATGCATGTA		6840
P K N F E E A I P E K A L L T G E L E E		
CCCCAAAATACTTTGAAGAGGCTATTCTCTGAGAAAGCTCTGTTAACTGGTGGCTAGAGA		6900
T N E G Y A I A K I A V A K A C E Y I S		
AACTATGAGGGATATGCTATTGCGAAAATTCCTGTAGCAAAAAGCATGCCAATATATATC		6960
R E N S N Y F Y K T I I P C N L Y G K Y		
AAGAGAAAATCTTAATTTATTTTATAAAACAATTTATCCCATGTAATTTATATGGGAATA		7020
D K F D D N S S H M I P A V I K K I H H		
TCATAAATTTGATGATAAATCTCTCACATATGATTCCCGCACTTATAAAAAAATCCATCA		7080
A K I N N V P E I E I W G D G N S R R E		
TGCGAAAATTAATAATGTCCTCCAGAGATCGAAATTTGGGGGGATGGTAATTCGCGCGCTGA		7140
F M Y A E D L A D L I F Y V I P K I E F		
GTTTATGTATGCAGAAAGATTTAGCTGATCTTATTTTATGTTATTCCTAAAATAGAAAT		7200
M P N M V N A G L G Y D Y S I N D Y Y K		
CATGCCATATATGCTAATGCTGGTTTAGCTTACGATTATTCGAATTAATGACTATTATAA		7260
I I A E E I G Y T G S F S H D L T K P T		
GATAATTCGAGAAAGAAATTCGTTATACTGGGAGTTTTTCTCATGATTTAACAAAACCAAC		7320
G M K R K L V D I S L L N K I G W S S H		
AGGAATGAAACGGAAAGCTAGTAGATATTTCAATTCCTTAATAAAATTCGTTGCTCAAGTCA		7380
F E L R D G I R K T Y N Y Y L E N Q N K		
CTTTGAACTCAGAGATGGCATTCAGAAAGACCTATAATTTATTACTTGGAGAAATCAAAATAA		7440

Start of orf7, End of orf6

M I T Y P L A S N T W D E Y E Y A A I Q	
*	
ATGATTACATACCCCACTTGGTAAATAAATCTGGGATGAATATGAGTATGCAGCAATACAG	7500
S V I D S K M F T M G K K V E L Y E K N	
TCAGTAATTCAGCTCAAAAATGTTTACCATGGGTAAAAGCTTGAATTTATGAGAAAAAT	7560
F A D L F G S K Y A V M V S S G S T A N	
TTTCTCATTTTCTTTGGTAGCAAAATATGCCCTAATGCTTAGCTCTGGTTCTACAGCTAAT	7620

Figure 5/6

L L M I A A L F F T N K P K L K R G D E CTGTTAATGATTGCTGCCCCCTTTCTTCACTAATAAACCAAACTTAAAGAGGTGATGAA	7680
I I V P A V S W S T T Y Y P L Q Q Y G L ATAATAGTACCTGCACTGTCATGGTCTACGACATATTACCCTCTGCAACAGTATGCGTTA	7740
K V K F V D I N K E T L N I D I D S L K AAGGTGAAGTTTGTGCGATATCAATAAAGAACTTTAAATATTGATATCGATAGTTTGAAA	7800
N A I S D K T K A I L T V N L L G N P N AATGCTATTTTCAGATAAAACAAAAGCAATATTGACAGTAAATTTATTAGGTAATCTTAAT	7860
D F A K I N E I I N N R D I I L L E D N GATTTTGCATAAATAATGAGATAATAAATAATAGGGATATTATCTTACTAGAGATAAC	7920
C E S M G A V F Q N K Q A G T F G V M G TGTGASTCGATGGGCGCGCTCTTCAAAATAAGCAGGCAAGGCACATTCGGAGTTATGGGT	7980
T F S S F Y S H H I A T M E G G C V V T ACCTTTAGTTCTTTTACTCTCATCATATAGCTACAAATGGAAGGGGGCTGCGTAGTTACT	8040
D D E E L Y H V L L C L R A H G W T R N GATGATGAGAGCTGTATCATGTATTGTTGTGCGCTTCGAGCTCATGTTGGACAAGAAAT	8100
L P K E N M V T G T K S D D I F E E S F TTACCAAAAGAGAAATATGCTTACAGGCACTAAGAGTGTGATATTTTGGAAAGCTCGTTT	8160
K F V L P G Y N V R P L E M S G A I G I AAGTTTGGTTTACCAGGATACAAATGTTGCGCCCACTTGAAATGAGTGGTGTATTTGGGATA	8220
E Q L K K L P G F I S T R R S N A Q Y F GAGCACTTAAAAAGTTTACCAGGTTTATATCCACCAGAGCTTCCAAATGCACAAATATTTT	8280
V D K F K D H P F L D I Q K E V G E S S GTAGATAAATTTAAGATCATCATTCCTTGATATACAAAAGAAAGTTGGTGAAAGTAGG	8340
W F G F S F V I K E G A A I E R K S L V TGCTTTGCTTTTTCCTTCGTTATAAAGGAGGGAGCTGCTATTGAGAGGAAGAGTTTAGTA	8400
N N L I S A G I E C R P I V T G N F L K AATAATCTGATCTCAGCAGGCATTGAAATGCGGACCAATGTTTACTGGCAATTTTCTCAA	8460
N E R V L S Y F D Y S V H D T V A N A E AATGAAGCTGTTTTGAGTTATTTTGATTACTCTGTACATGATACGGTAGCAATGCCGAA	8520
Y I D K N G F F V G N H Q I P L F N E I TATATAGATAAGAAATGGTTTTTTTGTGCGAAACCACAGATACCTTTGTTTAATGAATA	8580
End of orf7	
D Y L R K V L K *	
GATTATCTACGAAAGTATTAAAAAATACTAACGAGGCACTCTATTTCGAATAGAGTGCCT	8640
Start of orf8	
M V L T V K K I L A F G Y S K V L P TTAAGATGCTATTACAGTGAATAAAATTTAGCGTTTGGCTATTCTAAAGTACTACCAE	8700
P V I E Q F V N P I C I F I I T P L I L CGGTTATGAAACAGTTTGTCAATCCAAATTTGCATCTTCAATTATCACACCACTAATACTCA	8760
N H L G K Q S Y G N W I L L I T I V S F ACCACCTGGGTAAGCAAGCTATGCTAATTCGATTTTATTAATTACTATTGTATCTTTTT	8820

Figure 5/7

S Q L I C G G C S A W I A K I I A E Q R CTCAGTTAATATGTGGAGGATGTTCCGCATGGATTGCAAAATCATTGCAGAACAGAGAA	8880
I L S D L S K K N A L R Q I S Y N F S I TTCTTAACTGATTTATCAAAAAAATGCTTTTACGTCAAATTCCTATAATTTTTCAATTG	8940
V I I A F A V L I S F L I L S I C F F D TTATTATCGCATTTCGGGTATTGATTTCTTTTCTTATATTAAGTATTTCTTTCTTCGATG	9000
V A R N N S S F L F A I I I C G F F Q E TTCCGAGGAATAATTCTTCATTCTTATTCCGGATTATTTATTTGTGGTTTTTTTCAGGAAG	9060
V D N L F S G A L K G F E K F N V S C F TTGATAATTTATTTTACTGCTGGCTAAAGGTTTTGAAAAATTTAATGTATCATGTTTTT	9120
F E V I T R V L W A S I V I Y G I Y G N TTCAAGTAATTACAAGAGTGCTCTGGGCTTCTATAGTAATATATGGCATTTCACGGAATG	9180
A L L Y F T C L A F T I K G M L K Y I L CACTCTTATATTTTACATGTTTACGCTTTACCATTAAGGTATGCTAAATATATTTCTTG	9240
V C L N I T G C F I N P N F N R V G I V TATGCTGTGATATATTACGGTTGTTTCATCAATCCTAATTTAATAGAGTTGGGATTGTTA	9300
N L L N E S K W M F L Q L T G G V S L S ATTTGTTAATGAGTCAAATGGATGTTTTCTTCAATTAAGTGGTGGCGTCTCACTTACTT	9360
L F D R L V I P L I L S V S K L A S Y V TGTTTGATAGGCTCGTAATACCATTCATTTTTATCTGTCTAGTAAACTGGCTTCTTATCTCC	9420
P C L Q L A Q L M F T L S A S A N Q I L CTTGGCTTCAACTAGCTCAATTGATGTTCACTCTTTCTGGCTGTGCAATCAATATTAC	9480
L P M F A R M K A S N T F P S N C F F K TACCAATGTTTGGTAGAATGAAGCATCTAACACATTTCCTCTAATTTGTTTTTTTAAAA	9540
I L L V S L I S V L P C L A L F F F G R TTCTGCTGTATCACTAATTTCTGTTTTGCTTGTCTTGGCTTATTCTTTTTTGGTCTGTG	9600
D I L S I W I N P T F A T E N Y K L M Q ATATATTATCAATATGGAATAACCTACATTTCGCAACTGAAATATATAAATTAATGCAAA	9660
I L A I S Y I L L S M M T S F H F L L L TTTTAGCTATAAGTTACATTTTATTTGTCATGATGACATCTTTTCATTCTCTGTTATTAG	9720
G I G K S K L V A N L N L V A G L A L A GAATTGCTAATCTAAGCTTGTTCGCAATTTAATCTGTTGCAAGGCTTCGCACTTCTG	9780
A S T L I A A H Y G L Y A I S M V K I I CTTCAAGCTTAATCGCAGCTCATTTATGGCTTTATGCAATATCTATGGTAAAAATATAT	9840
Y P A F Q F Y Y L Y V A F V Y F N R A K ATCCGGCTTTTCAATTTTATTACCTTTATGTAGCTTTTGTCTATTTTAAATAGACGGAAA	9900

Start of orf9, End of orf8

M S I D L L F S I T E I A I V F S C T I N V Y *	
ATGCTCTATGATTTACTTTTTTCAATTAAGTGAATCGCAATGTTTTTTCTTGGCACTATT	9960
Y I F T Q C L L M R R I Y L D K S I L I TACATATTTTACTCAATGTTTGTAAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATT	10020
L L C L L F F L V I I Q L P E L N V N G CTTTTATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGT	10080

Figure 5/8

L V D S L K L S L P L L M V F I A F Q K TTGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAA	10140
P K L C L W V I I A L L F L N S A F N F CCGAAATTATGCTTGTGGGTATTATTGCATTGTTGTTTTGAACCTGCATTAAATTTT	10200
L Y L K T F D K F S S F P F T F F I L L TTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCCTTTTACTTTTTTTATATTGCTG	10260
F Y L F R L G I G N L P V Y K N K K F Y TTTTACTTGTTTAGATTGGGAATTGGTAATTACCGGTTTATAAAAAATAAAAAATTTTAC	10320
A L I F L F I L I D I M Q S L L I N Y R GCGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGG	10380
G Q I L Y S V I C I L I L V F K V N L R GGGCAGATTTTATATCCGTAATTGTCATCCTGATACTTGTGTTTAAAGTTAATTAAAGA	10440
K K I P Y F F L M L P V L Y V I I M A Y AAAAAGATTCCATACATTTTTTTTAAATGCTGCCAGTTTTTATATGTAATTATTATGGCTTAT	10500
I G F N Y F N K G V T F F E P T A S N I ATTGGTTTTAATTATTCAATAAAGCGTAACTTTTTTTGAACCTACAGCAAGTAATATT	10560
E R T G M I Y Y L V S Q L G D Y I F H G GAACGTACGGGGATGATATATTATTGTTTCACAGCTTGGTGATTATATATTCCATGGT	10620
M G T L N F L N N G G Q Y K T L Y G L P ATGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCA	10680
S L I P N D P H D F L L R F F I S I G V TCATTAATTCCTAATGACCCTCATGATTTTTTTATTACGGTTCTTTATAAGTATTGGTGTG	10740
I G A L V Y H S I F F V F F R R I S F L ATAGGAGCATTTGGTTTATCATTCTATATTTTTTGTTTTTTTTAGGAGAATATCTTCTTA	10800
L Y E R N A P F I V V S C L L L L Q V V TTATATGAGAGAAATGCTCCTTTCATGTTGTAAGTTGTTTGTACTGTACAAGTTGTG	10860
L I Y T L N P F D A F N R L I C G L T V TTAATTTATACATTAAACCCTTTTGATGCTTTTAAATCGATTGATTTCGGGGCTTACAGTT	10920
 Start of orf10 End of orf9	
G V V Y G F A K I R *	
M D L Q K L D K Y T C N G N L D A GGAGTTGTTTATGGATTGCAAAAATTAGATAAGTATACCTGTAATGGAAATTTAGACGC	10980
P L V S I I I A T Y N S E L D I A K C L TCCACTTGTTTCAATAATCATTGCAACTTATAATTCTGAACTTGATATAGCTAAGTGT	11040
Q S V T N Q S Y K N I E I I I M D G G S GCAATCGGTAACATAATCAATCTTATAAGAATATTGAAATCATAATAATGGATGGAGGATC	11100
S D K T L D I A K S F K D D R I K I V S TTCTGATAAAACGCTTGATATTGCAAAATCGTTTAAAGACGACCGAATAAAAAATAGTTTC	11160
E K D R G I Y D A W N K A V D L S I G D AGAGAAAGATCGTGGAAATTATGATGCCTGGAATAAAGCAGTTGATTATCCATTGGTGA	11220
W V A F I G S D D V Y Y H T D A I A S L TTGGGTAGCATTTATTGGTTCAGATGATGTTTACTATCATACAGATGCAATTGCTTCATT	11280
M K G V M V S N G A P V V Y G R T A H E GATGAAGGGGGTTATGGTATCTAATGCGCCCCGTGGTTTATGGGAGGACAGCGCACGA	11340

Figure 5/9

G P D R N I S G F S G S E W Y N L T G F AGGTCCCAGATAGGAACATATCTGGATTTTCAGGCAGTGAATGGTACAACCTAACAGGATT	11400
K F N Y Y K C N L P L P I M S A I Y S R TAAGTTTAATTATTACAAATGTAATTTACCATTGCCCATTTATGAGCGCAATATATTCTCG	11460
D F F R N E R F D I K L K I V A D A D W TGATTTCTTCAGAAACGAACGTTTGTATATTAAATTAAAAATTGTTGCTGACGCTGATTG	11520
F L R C F I K W S K E K S P Y F I N D T GTTTCTGAGATGTTTCATCAAATGGAGTAAAGAGAAGTCACCTTATTTTATTAATGACAC	11580
T P I V R M G Y G G V S T D I S S Q V K GACCCCTATGTGTTAGATGGGATATGGTGGGGTTTCGACTGATATTTCTTCTCAAGTTAA	11640
T T L E S F I V R K K N N I S C L N I Q AACTACGCTAGAAAGTTTCATTGTACGCAAAAAGAATAATATATCCTGTTTAAACATACA	11700
L I L R Y A K I L V M V A I K N I F G N GCTGATTCCTAGATATGCTAAAAATCTGGTGATGGTAGCGATCAAAAATATTTTGGCAA	11760
N V Y K L M H N G Y H S L K K I K N K I TAATGTTTATAAATTAATGCATAACGGGTATCATTCCTTAAAGAAAAATCAAGAATAAAAT	11820
Start of orf11, End of orf10	
M K I V Y I I T G L T C G G A E H L M T *	
<u>ATGAAGATGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG</u>	11880
Q L A D Q M F I R G H D V N I I C L T G CAGTTAGCAGACCAATGTTTATACGCGGCATGATGTTAATATTATTTGTCTAACTGGT	11940
I S E V K P T Q N I N I H Y V N M D K N ATATCTGAGGTAAAGCCAACACAAAATATTAATATTTCATTATGTTAATATGGATAAAAAAT	12000
F R S F F R A L F Q V K K I I V A L K P TTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAATAATTGTGCGCTTAAAGCCA	12060
D I I H S H M F H A N I F S R F I R M L GATATAATACATAGTCATATGTTTCATGCTAATATTTTATGTCGTTTATTAGGATGCTG	12120
I P A V P L I C T A H N K N E G G N A R ATTCCAGCGGTGCCCCGTGATATGTACCGCACACAACAAAATGAAGTGCGCAATGCAAGG	12180
M F C Y R L S D F L A S I T T N V S K E ATGTTTGTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGTAAAGAG	12240
A V Q E F I A R K A T P K N K I V E I P GCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGATTCCG	12300
N F I N T N K F D F D I N V R K K T R D AATTTTATTAATACAAATAAAATTTGATTTTGATATTAAATGTCAGAAAGAAAACGCGAGAT	12360
A F N L K D S T A V L L A V G R L V E A GCTTTTAAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTGAAGCA	12420
K D Y P N L L N A I N H L I L S K T S N AAAGACTATCCGAACCTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACATCAAAT	12480
C N D F I L L I A G D G A L R N K L L D TGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGAT	12540
L V C Q L N L V D K V F F L G Q R S D I TTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATT	12600

Figure 5/10

K E L M C A A D L F V L S S E W E G F G AAAGAATTAATGTGTGCTGCAGATCTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTGGT	12660
L V V A E A M A C E R P V V A T D S G G CTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGA	12720
V K E V V G P H N D V I P V S N H I L L GTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTG	12780
A E K I A E T L K I D D N A R K I I G M GCAGAGAAAATCGCTGAGACACTTAAATAGATGATAACGCAAGAAAAATAAGGTATG	12840
K N R E Y I V S N F S I K T I V S E W E AAAAATAGAGAATATATTGTTTCCAATTTTCAATTAACGATAGTGAGTGAGTGGGAG	12900
End of orf11	
R L Y F K Y S K R N N I I D * CGCTTATATTTTAAATATTCCAAGCGTAATAATATAATTGAT TGAAAATATAAGTTTGTA	12960
CTCTGGATGCAATAGTTTCTCTATGCTGTTTTTTTACTGGCTCCGTATTTTTTACTTATAG	13020
CTGGATTTTGTATATATCAGTATTAATCTGTCTCAACTTCATCTAGACTACATTCAAGC	13080
Start of gnd	
M S K Q Q I CGCGCATGCGTCGCGCGGTGACTACACCTGACAGGAGTATGTA ATGTCCAAGCAACAGAT	13140
G V V G M A V M G R N L A L N I E S R G CGGCGTCGTGCTGATGGGCGCAACCTGGCGCTCAACATCGAAAGCCGCGG	13200
Y T V S I F N R S R E K T E E V V A E N TTATACCGTCTCCATCTTCAACCGCTCCCGCGAGAAAACGAAGAAGTTGTTGCCGAGAA	13260
P D K K L V P Y Y T V K E F V E S L E T CCCGGATAAGAACTGGTTCCTTATTACACGGTGAAGAGTTCGTCGAGTCTCTTGAAAC	13320
P R R I L L M V K A G A G T D A A I D S CCCACGTCGTATCTGTTAATGGTAAAAGCAGGGGCGGGAAGTATGCTGCTATCGATTC	13380
L K P Y L D K G D I I I D G G N T F F Q CCTGAAGCCGTATCTGGATAAAGGCACATCATATTGATGGTGGCAACACCTTCTTCCA	13440
D T I R R N R E L S A E G F N F I G T G GGACACTATCCGTCGTAACCGTGAACGTGCCGGAAGGCTTTAACTTCATCGGTACCGG	13500
V S G G E E G A L K G P S I M P G G Q K CGTGTCCGGCGGTGAAGAGGGCGCCCTGAAAGGCCATCTATCATGCCAGGTGGCCAGAA	13560
E A Y E L V A P I L T K I A A V A E D G AGAAGCGTATGAGCTGGTTGCGCCTATCCTGACCAAGATTGCTGCGGTTGCTGAAGATGG	13620
E P C I T Y I G A D G A G H Y V K M V H CGAACCATGTATAACTTACATCGGTGCTGACGGTGCGGGTCACTACGTGAAGATGGTGCA	13680
N G I E Y G D M Q L I A E A Y S L L K G CAACGGTATCGAATATGGCGATATGCAGCTGATTGCTGAAGCCTATTCTCTGCTTAAAGG	13740
G L N L S N E E L A T T F T E W N E G E CGGCCTTAATCTGTCTAACGAAGAGCTGGCAACCACTTTTACCGAGTGGAATGAAGGCGA	13800
L S S Y L I D I T K D I F T K K D E E G GCTAAGTAGCTACCTGATTGACATCACCAAAGACATCTTCACCAAAAAGATGAAGAGGG	13860

Figure 5/11

K Y L V D V I L D E A A N K G T G K W T TAAATACCTGGTTGATGTGATCCTGGACGAAGCTGCGAACAAAGGCACCGGTAAATGGAC	13920
S Q S S L D L G E P L S L I T E S V F A CAGCCAGAGCTCTCTGGATCTGGGTGAACCGCTGTCGCTGATCACCGAATCCGTATTTCGC	13980
R Y I S S L K D Q R I A A S K V L S G P TCGCTACATCTCTTCTCTGAAAGACCAGCGCATTGCGGCATCTAAAGTGCTGTCTGGTCC	14040
Q A K L A G D K A E F V E K V R R A L Y GCAGGCTAAACTGGCTGGTGATAAAGCAGAGTTTCGTTGAGAAAGTCCGTCGCGCGCTGTA	14100
L G K I V S Y A Q G F S Q L R A A S D E CCTGGGTAAAATCGTCTCTTATGCCCAAGGCTTCTCTCAACTGCGTGCCGCGTCTGACGA	14160
Y N W D L N Y G E I A K I F R A G C I I ATACAACCTGGGATCTGAACTACGGCGAAATCGCGAAGATCTTCCGCGCGGGCTGCATCAT	14220
R A Q F L Q K I T D A Y A E N K G I A N TCGTGCGCAGTTCCTGCAGAAAATTACTGACGCGTATGCTGAAAACAAAGGCATTGCTAA	14280
L L L A P Y F K N I A D E Y Q Q A L R D CCTGTTGCTGGCTCCGTACTTCAAAAATATCGCTGATGAATATCAGCAAGCGCTGCGTGA	14340
V V A Y A V Q N G I P V P T F S A A V A TGTAAGTGGCTTATGCTGTGCAGAACGGTATTCGGGTACCGACCTTCTCTGCAGCGGTAGC	14400
Y Y D S Y R S A V L P A N L I Q A Q R D CTACTACGACAGCTACCGTTCTGCGGTACTGCCGGCTAATCTGATTTCAGGCACAGCGTGA	14460
Y F G A H T Y K R T D K E G V F H T G TTACTTCGGTGCGCACACGTATAAACGCACTGATAAAGAAGGTGTGTTCCACACCG	14516

Figure 5/12

GTAACCAAGGGCGGTACGTGCATAAATTTTAATGCTTATCAAAACTATTAGCATTAAAAA 60

Start of orf1
M N K E T V S I I M P V Y N

TATATAAGAAATTCCTCAAATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAAT 120

G A K T I I S S V E S I I H Q S Y Q D F
GGGGCCAAACTATAATCTCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTT 180

V L Y I I D D C S T D D T F S L I N S R
GTTTTGTATATCATTGACGATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGA 240

Y K N N Q K I R I L R N K T N L G V A E
TACAAAAACAATCAGAAAAATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAA 300

S R N Y G I E M A T G K Y I S F C D A D
AGTCGAAATATGGAATAGAAATGGCCACGGGAAATATATTTCTTTTGTGATGCGGAT 360

D L W H E K K L E R Q I E V L N N E C V
GATTTGTGGCAGAGAAAAAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTA 420

D V V C S N Y Y V I D N N R N I V G E V
GATGTGGTATGTTCTAATTATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTT 480

N A P H V I N Y R K M L M K N Y I G N L
AATGCTCCTCATGTGATAAATTATAGAAAAATGCTCATGAAAACTACATAGGGAATTTG 540

T G I Y N A N K L G K F Y Q K K I G H E
ACAGGAATCTATAATGCCAACAAATTGGGTAAGTTTATCAAAAAAGATTGGTCACGAG 600

D Y L M W L E I I N K T N G A I C I Q D
GATTATTTGATGTGGCTGGAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGAT 660

N L A Y Y M R S N N S L S G N K I K A A
AATCTGGCGTATTACATGCGTTCAAATAATTCATATCGGGTAATAAAATTAAAGCTGCA 720

K W T W S I Y R E H L H L S F P K T L Y
AAATGGACATGGAGTATATATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATAT 780

Y F L L Y A S N G V M K K I T H S L L R
TATTTTTTATATATGCTTCAAATGGAGTCATGAAAAAATAACACATTCATATTAAGG 840

Start of orf2, End of orf1
R K E T K K *

V K S A A K L I F L F L F T
AGAAAGGAGACTAAAAAGTGAAGTCAGCGGCTAAGTTGATTTTATTCCTATTTACAC 900

L Y S L Q L Y G V I I D D R I T N F D T
TTTATAGTCTCCAGTTGTATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAA 960

K V L T S I I I I F Q I F F V L L F Y L
AGGTATTAACTAGTATTATAATTATATTTTCAGATTTTTTTTGTGTTTATTATTTATCTAA 1020

T I I N E R K Q Q K K F I V N W E L K L
CGATTATAAATGAAAGAAAACAGCAGAAAAAATTTATCGTGAAGTGGGAGCTAAAGTTAA 1080

I L V F L F V T I E I A A V V L F L K E
TACTCGTTTTCTTTTGTGACTATAGAAATGCTGCTGTAGTTTTTATTTCTTAAAGAAG 1140

G I P I F D D D P G G A K L R I A E G N
GTATTCCTATATTTGATGATGATCCAGGGGGGCTAAACTTAGAATAGCTGAAGGTAATG 1200

Figure 6/1

G L Y I R Y I K Y F G N I V V F A L I I GACTTTACATTAGATATATTAAGTATTTTGGTAATATAGTTGTGTTGCATTAATTATTC	1260
L Y D E H K F K Q R T I I F V Y F T T I TTTATGATGAGCATAAATTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTG	1320
A L F G Y R S E L V L L I L Q Y I L I T CTTTATTTGGTTATCGTTCTGAATTGGTGTGCTCATTCTTCAATATATATTGATTACCA	1380
N I L S K D N R N P K I K R I I G Y F L ATATCCTGTCAAAGGATAACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTAT	1440
L V G V V C S L F Y L S L G Q D G E Q N TGGTAGGGGTGTATGCTCGTTGTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATG	1500
D S Y N N M L R I I N R L T I E Q V E G ACTCATATAATAATATGTTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTG	1560
V P Y V V S E S I K N D F F P T P E L E TTCCATATGTTGTTTCTGAATCTATTAAGAACGATTTCCTTCCGACACCAGAGTTAGAAA	1620
K E L K A I I N R I Q G I K H Q D L F Y AGGAATTTAAAGCAATAATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATG	1680
G E R L H K Q V F G D M G A N F L S V T GAGAACGGTTACATAAACAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTA	1740
T Y G A E L L V F F G F L C V F I I P L CGTATGGAGCAGAACTGTAGTTTTTTTTGGTTTTCTCTGTGTATTTCATTATCCCTTTAG	1800
G I Y I P F Y L L K R M K K T H S S I N GGATATATATACCTTTTTATCTTTTAAAGAGAATGAAAAAACCCATAGCTCGATAAATT	1860
C A F Y S Y I I M I L L Q Y L V A G N A GCGCATTCATTATCATATATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCAT	1920
S A F F F G P F L S V L I M C T P L I L CGGCCTTCTTTTTTGGTCCTTTCTCTCCGTATTGATAATGTGTACTCTCTGATCTTAT	1980
Start of orf3	
M K I S V I T V T Y L H D T L K R L S R N E N I S Y N C D L TGCATGATACGTTAAAGAGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTAT	2040
End of orf2	
N N A E G L E K T L S S L S I L K I K P AATAATGCTGAAGGGTTAGAAAAACTTTAAGTAGTTTATCAATTTTAAAAATAAAACCT	2100
F E I I I V D G G S T D G T N R V I S R TTTGAGATTATTATAGTTGATGGCGGCTCTACAGATGGAACGAATCGTGTCAATTAGTAGA	2160
F T S M N I T H V Y E K D E G I Y D A M TTTACTAGTATGAATATTACACATGTTTATGAAAAAGATGAAGGGATATATGATGCGATG	2220
N K G R M L A K G D L I H Y L N A G D S AATAAGGGCCGAATGTTGGCCAAAGGCGACTTAATACATTATTTAAACGCCGCGCATAGC	2280
V I G D I Y K N I K E P C L I K V G L F GTAATTGGAGATATATATAAAATATCAAAGAGCCATGTTTGATTAAAGTTGGCCTTTTC	2340
E N D K L L G F S S I T H S N T G Y C H GAAAATGATAAACTTCTGGGATTTTCTTCTATAACCCATTCAAATACAGGGTATGTGCAT	2400

Figure 6/2

Q G V I F P K N H S E Y D L R Y K I C A CAAGGGGTGATTTTCCCAAAGAATCATTTCAGAAATATGATCTAAGGTATAAAATATGTGCT	2460
D Y K L I Q E V F P E G L R S L S L I T GATTATAAGCTTATTCAAGAGGTGTTTCCTGAAGGGTTAAGATCTCTATCTTTGATTACT	2520
S G Y V K Y D M G G V S S K K R I L R D TCGGGTATGTAAAAATATGATATGGGGGAGTATCTTCAAAAAAAGAATTTTAAGAGAT	2580
K E L A K I M F E K N K K N L I K F I P AAAGAGCTTGCCAAAATTATGTTTGAAAAAATAAAAAAACCTTATTAAGTTTATTCCA	2640
I S I I K I L F P E R L R R V L R K M Q ATTTCAATAATCAAAATTTTATTCCTGAACGTTTAAGAAGAGTATGCGGAAAATGCAA	2700
<div style="text-align: center;">Start of orf4 End of orf3</div>	
Y I C L T L F F M K N S S P Y D N E * M I M N K I TATATTTGTCTAACTTTATTCTTCATGAAGAATAGTTCACCATATGATAATGAATAAAAT	2760
K K I L K F C T L K K Y D T S S A L G R CAAAAAATACTTAAATTTTGCACTTTAAAAAATATGATACATCAAGTGCTTTAGGTAG	2820
E Q E R Y R I I S L S V I S S L I S K I AGAACAGGAAAGGTACAGGATTATATCCTTGCTGTATTTCAGTTTGATTAGTAAAT	2880
L S L L S L I L T V S L T L P Y L G Q E ACTCTCACTACTTTCTCTTATATTAAGTGAAGTTTAACTTTACCTTATTTAGGACAAGA	2940
R F G V W M T I T S L G A A L T F L D L GAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCTCTGACATTTTGGACTT	3000
G I G N A L T N R I A H S F A C G K N L AGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCAATTTGCGTGTGGCAAAAATTT	3060
K M S R Q I S G G L T L L A G L S F V I AAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCTGGATTATCGTTTGTCTAT	3120
T A I C Y I T S G M I D W Q L V I K G I AACTGCAATATGCATATTACTTCTGGCATGATTGATTGGCAACTAGTAATAAAAGGTAT	3180
N E N V Y A E L Q H S I K V F V I I F G AAACGAGAATGTGTATGCAGAGTTACAACACTCAATTAAAGTCTTTGTAATCATATTTGG	3240
L G I Y S N G V Q K V Y M G I Q K A Y I ACTTGGAATTTATTCAAATGGTGTGCAAAAAGTTTATATGGGAATACAAAAGCCTATAT	3300
S N I V N A I F I L L S I I T L V I S S AAGTAATATTGTTAATGCCATATTTATATTGTTATCTATTACTCTAGTAATATCGTC	3360
K L H A G L P V L I V S T L G I Q Y I S GAACTACATGCGGGACTACCAGTTTAAATGTGCAGCACTCTTGGTATTCAATACATATC	3420
G I Y L T I N L I I K R L I K F T K V N GGGAATCTATTTAAACAATTAATCTTATTATAAAGCGATTAAATAAAGTTTACAAAAGTTAA	3480
I H A K R E A P Y L I L N G F F F F I L CATACATGCTAAAAGAGAAGCTCCATATTTGATATTAAACGGTTTTTCTTTTTTATTTT	3540
Q L G T L A T W S G D N F I I S I T L G ACAGTTAGGCACTCTGGCAACATGGAGTGGTGATAACTTTTATAATATCTATAACATTGGG	3600

Figure 6/3

V T Y V A V F S I T Q R L F Q I S T V P TGTTACTTATGTTGCTGTTTTTAGCATTACACAGAGATTATTTCAAATATCTACGGTCCC	3660
L T I Y N I P L W A A Y A D A H A R N D TCTTACGATTTATAACATCCCGTTATGGGCTGCTTATGCAGATGCTCATGCACGCAATGA	3720
T Q F I K K T L R T S L K I V G I S S F TACTCAATTTATAAAAAAGACGCTCAGAACATCATTGAAAATAGTGGGTATTTTCATCATT	3780
L L A F I L V V F G S E V V N I W T E G CTTATTGGCCTTCATATTAGTAGTGTTCGGTAGTGAAGTCGTTAATATTTGGACAGAAGG	3840
K I Q V P R T F I I A Y A L W S V I D A AAAGATTACAGGTACCTCGAACATTCATAATAGCTTATGCTTTATGGTCTGTTATTGATGC	3900
F S N T F A S F L N G L N I V K Q Q M L TTTTTCGAATACATTTGCAAGCTTTTAAATGGTTTGACATAGTTAAACAACAAATGCT	3960
A V V T L I L I A I P A K Y I I V S H F TGCTGTTGTAACATTGATATTGATCGCAATTCCAGCAAAATACATCATAGTTAGCCATTT	4020
G L T V M L Y C F I F I Y I V N Y F I W TGGGTTAACTGTTATGTTGTAAGTCTTCATTTTATATATATTGTAATTAATTAATTAATG	4080
Start of orf5, End of orf4 M K M	
Y K C S F K K H I D R Q L N I R G * GTATAAATGTAAGTTTAAAAACATATCGATAGACAGTTAAATATAAGAGGATGAAAATG	4140
K Y I P V Y Q P S L T G K E K E Y V N E AAATATATACCAAGTTTACCAACCGTCATTGACAGGAAAAGAAAAGAAATATGTAATGAA	4200
C L D S T W I S S K G N Y I Q K F E N K TGTCTGGAAGTCAACGTCGATTTCAATCAAAAGGAAACTATATTCAGAAAGTTTGAAAATAAA	4260
F A E Q N H V Q Y A T T V S N G T V A L TTTGCGGAACAAAACCATGTCGAATATGCAACTACTGTAAGTAATGGAACGGTTGCTCTT	4320
H L A L L A L G I S E G D E V I V P T L CAATTAAGCTTTGTTAGCGTTAGGTATATGCGAAGGAGATGAAGTTATGTTTCCAAAGT	4380
T Y I A S V N A I K Y T G A T P I F V D ACATATATAGCATCAAGTTAATGCTATAAATACACAGGAGCCACCCCATTTTCTGTTGAT	4440
S D N E T W Q M S V S D I E Q K I T N K TCAGATATGAAGCTTGGCAAAATGTTGTTAGTGAATAGAACAAAATCACTATATAA	4500
T K A I M C V H L Y G H P C D M E Q I V ACTAAAGCTATTATGTTGTTGCTCAATTTATACGGACATCCATGTCATATGGAACAAATGTA	4560
E L A K S R N L F V I E D C A E A F G S GAAGTGGCCAAAAGTACAAATTTGTTTGAATTTGAAGATTGCGCTGAAGCCCTTTGGTTCT	4620
K Y K G K Y V G T F G D I S T F S F F G AAATATAAGGTAATATGTTGGGACATTTGGAGATATTTCTACTTTTAGCTTTTTTGGAA	4680
N K T I T T G E G G M V V T N D K T L Y AATAAACTATTACTACAGGTTAAGGTGGAATGCTTGTCAAGAAATGACAAAACACTTTAT	4740
D R C L H F K G Q G L A V H R Q Y W H D GACCGTTGTTTACATTTTAAAGGCCAAGGATTAGCTGTACATAGGCCAATATTGGCATGAC	4800
V I G Y N Y R M T N I C A A I G L A Q L GTTATAGGCTACAAATATAGGATGACAAATATCTGCGCTGCTATAGGATTAGCCCACTTA	4860

Figure 6/4

E Q A D D F I S R K R E I A D I Y K K N	
GAACAAGCTGATGATTTTATATCAGGAAAACGTGAAATTGCTGATATTTATAAAAAAAT	4920
I N S L V Q V H K E S K D V F H T Y W M	
ATCAACAGTCTTTGTACAACTCCACAAGGAAAGTAAAGATGTTTTTCACACTTATTGCGATG	4980
V S I L T R T A E E R E E L R N H L A D	
GTCTCAATTCTTAAGTAGGACCCAGAGGAAAGAGAGGAATTAAGGAATCACCTTGCAGAT	5040
K L I E T R P V F Y P V H T M P M Y S E	
AAATCATCGAAACAAGGCCAATTTTTTACCTGTCCACAGGATGCCAATGTACTCGGAA	5100
K Y Q K H P I A E D L G W R G I N L P S	
AAATATCAAAAGCACCTATAGCTGAGGATCTTGGCTTGGCGTGGAAATTAATTTACCTAGT	5160
F P S L S N E Q V I Y I C E S I N E F Y	
TTCCCCAGCCTATCGAATGAGCAAGTTATTATATTTTGTGAATCTATTAACGAATTTTAT	5220
End of orf5	Start of orf6
S D K *	M K I A L N S D
AGTGATAAATAGCCTAAATATTTTAAAGGTCATTCATGAAATTTGCGTTGAATTCAGAT	5280
G F Y E W G G G I D F I K Y I L S I L E	
GGATTTTACGAGTGGGGCGGTGGAAATGATTTTTATTAATATATTCTGTCAATATTAGAA	5340
T K P E I C I D I L L P R N D I H S L I	
ACGAAACCAGAAATATGTATCGATATTCTTTTACCGAGAAATGATATACATTCTCTTATA	5400
R E K A F P F K S I L K A I L K R E R P	
AGAGAAAAAGCATTTCCTTTTAAAGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCT	5460
R W I S L N R F N E Q Y Y R D A F T Q N	
CGATGGATTTCATTAAATAGATTTAATGAGCAATACTATAGAGATGCCTTTACACAAAAT	5520
N I E T N L T F I K S K S S A F Y S Y F	
AATATAGAGACGAATCTTACCTTTTATTAAGTAAGAGCTCTGCCTTTTATTATATTTT	5580
D S S D C D V I L P C M R V P S G N L N	
GATAGTAGCGATTGTGATGTTATTCTTCCTTGCGTGTTCCTTCGGGAAATTTGAAT	5640
K K A W I G Y I Y D F Q H C Y Y P S F F	
AAAAAAGCATGGATTGGTTATATTATGACTTTCAACACTGTACTATCCTTCATTTTTT	5700
S K R E I D Q R N V F F K L M L N C A N	
AGTAAGCGAGAAATAGATCAAAGGAATGTGTTTTTTAAATGATGCTCAATTGCGCTAAC	5760
N I I V N A H S V I T D A N K Y V G N Y	
AATATTATTGTTAATGCACATTCAGTTATTACCGATGCAAAATAATATGTTGGGAATTAT	5820
S A K L H S L P F S P C P Q L K W F A D	
TCTGCAAAACTACATTCTCTTCCATTAGTCCATGCCCTCAATTAATGGTTTCGCTGAT	5880
Y S G N I A K Y N I D K D Y F I I C N Q	
TACTCTGGTAATATTGCCAAATATAATATTGACAAGGATTATTTTATAATTTGCAATCAA	5940
F W K H K D H A T A F R A F K I Y T E Y	
TTTGGAAACATAAAGATCATGCAACTGCTTTTAGGGCATTAAATTTTACTGAATAT	6000
N P D V Y L V C T G A T Q D Y R F P G Y	
AATCCTGATGTTTATTTAGTATGCACGGGAGCTACTCAAGATTATCGATTCCCTGGATAT	6060
F N E L M V L A K K L G I E S K I K I L	
TTTAATGAATTGATGGTTTGGCAAAAAAGCTCGGAATTGAATCGAAAAATTAAGATATTA	6120

Figure 6/5

G H I P K L E Q I E L I K N C I A V I Q GGGCATATACCTAAACTTGAACAAATTGAATTAATCAAAAATTGCATTGCTGTAATACAA	6180
P T L F E G G P G G G V T F D A I A L G CCAACCTTATTTGAAGCGGGCCTGGAGGGGGGTAACATTTGACGCTATTGCATTAGGG	6240
K K V I L S D I D V N K E V N C G D V Y AAAAAAGTTATACTATCTGACATAGATGTCAATAAAGAAGTTAATTGCGGTGATGTATAT	6300
F F Q A K N H Y S L N D A M V K A D E S TTCTTTCAGGCCAAAAACCATTATTCATTAAATGACGCGATGGTAAAAGCTGATGAATCT	6360
K I F Y E P T T L I E L G L K R R N A C AAAATTTTATGAACCTACAACCTCTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGT	6420
End of orf6	
A D F L L D V V K Q E I E S R S * CGAGATTTTCTTTTAGATGTTGTGAAACAAGAAATTGAATCCCGATCTTAATATATTCAA	6480
Start of orf7	
M T K V A L I T G V T G Q D G S Y GAGGTATATAATGACTAAAGTCGCTCTTATTACAGGTGTAAGTGGACAAGATGGATCTTA	6540
L A E F L L D K G Y E V H G I K R R A S TCTAGCTGAGTTTTTGTCTTGATAAAGGGTATGAAGTTCATGGTATCAAACGCCGAGCCTC	6600
S F N T E R I D H I Y Q D P H G S N P N ATCTTTTAATACAGAACGCATAGACCATATTTATCAAGATCCACATGGTTCTAACCCTAA	6660
F H L H Y G D L T D S S N L T R I L K E TTTTCACITGCACTATGGAGATCTGACTGATTCATCTAACCTCACTAGAATTCTAAAGGA	6720
V Q P D E V Y N L A A M S H V A V S F E GGTACAGCCAGATGAAGTATATAATTTAGCTGCTATGAGTCACGTAGCAGTTTCTTTTGA	6780
S P E Y T A D V D A I G T L R L L E A I GTCTCCAGAAATATACAGCCGATGTGATGCAATTGGTACATTACGTTTACTGGAAGCAAT	6840
R F L G L E N K T R F Y Q A S T S E L Y TCGCTTTTTAGGATTGGAAAACAAAACGCGTTTCTATCAAGCTTCAACCTCAGAATTATA	6900
G L V Q E I P Q K E S T P F Y P R S P Y TGGACTTGTTCAGGAAATCCCTCAAAAAGAATCCACCCCTTTTATCCTCGTTCCCTTA	6960
A V A K L Y A Y W I T V N Y R E S Y G I TGCAGTTGCAAACTTTACGCATATTGGATCACGGTAAATTATCGAGAGTCATATGGTAT	7020
Y A C N G I L F N H E S P R R G E T F V TTATGCATGTAATGGTATATTGTTCAATCATGAATCTCCACGCCGTGGAGAAACGTTTGT	7080
T R K I T R G L A N I A Q G L E S C L Y AACAAGGAAAATTACTCGAGGACTTGCAAATATTGCACAAGGCTTGGAATCATGTTTGT	7140
L G N M D S L R D W G H A K D Y V R M Q TTTAGGGAATATGGATTTCGTTACGAGATTGGGGACATGCAAAAGATTATGTTAGAAATGCA	7200
W L M L Q Q E Q P E D F V I A T G V Q Y ATGGTTGATGTTACAACAGGAGCAACCCGAAGATTTTGTGATTGCAACAGGAGTCCAATA	7260
S V R Q F V E M A A A Q L G I K M S F V CTCAGTCCGTCAGTTTGTGCAAAATGGCAGCAGCACAACTTGGTATTAAGATGAGCTTTGT	7320

Figure 6/6

G K G I E E K G I V D S V E G Q D A P G	
TGGTAAAGGAATCGAAGAAAAAGGCATTGTAGATTCGGTTGAAGGACAGGATGCTCCAGG	7380
V K P G D V I V A V D P R Y F R P A E V	
TGTGAAACCAGGTGATGTCTATTGTTGCTGTTGATCCTCGTTATTTCCGACCAGCTGAAGT	7440
D T L L G D P S K A N L K L G W R P E I	
TGATACTTTGCTTGGAGATCCGAGCAAAGCTAATCTCAAACCTGGTTGGAGACCAGAAAT	7500
T L A E M I S E M V A K D L E A A K K H	
TACTCTTGCTGAAATGATTCTGAAATGGTTGCCAAAGATCTTGAAGCCGCTAAAAACA	7560
Start of orf8, End of orf7	
	M M M N K
S L L K S H G F S V S L A L E *	
TTCTCTTTTAAATCGCATGGTTTTCTGTAAAGCTTAGCTCTGGAATGATGATGAATAAG	7620
Q R I F I A G H Q G M V G S A I T R R L	
CAACGTATTTTATTGCTGGTCACCAAGGAATGGTTGGATCAGCTATTACCCGACGCCCTC	7680
K Q R D D V E L V L R T R D E L N L L D	
AAACAACGTGATGATGTTGAGTTGGTTTACGTACTCGGGATGAATTGAACCTGTTGGAT	7740
S S A V L D F F S S Q K I D Q V Y L A A	
AGTAGCGCTGTTTGGATTTTTTCTTCACAGAAAATCGACCAGGTTTATTTGGCAGCA	7800
A K V G G I L A N S S Y P A D F I Y E N	
GCAAAAGTCGGAGGTATTTTAGCTAACAGTTCTTATCCTGCCGATTTTATATATGAGAAT	7860
I M I E A N V I H A A H K N N V N K L L	
ATAATGATAGAGGCGAATGTCATTTCATGCTGCCCAAAAATAATGTAAATAAACTGCTT	7920
F L G S S C I Y P K L A H Q P I M E D E	
TTCTCGGTTGCTCGTGTATTATCTAAGTTAGCACACCAACCGATTATGGAAGACGAA	7980
L L Q G K L E P T N E P Y A I A K I A G	
TTATTACAAGGAACTTGAGCCAACAAATGAACCTTATGCTATCGCAAAAATTGCAGGT	8040
I K L C E S Y N R Q F G R D Y R S V M P	
ATTAAATTATGTGAATCTTATAACCGTCAGTTTGGGCGTGATTACCGTTCAGTAATGCCA	8100
T N L Y G P N D N F H P S N S H V I P A	
ACCAATCTTTATGGTCCAAATGACAAATTTTCATCCAAGTAATCTCATGTGATTCCGGCG	8160
L L R R F H D A V E N N S P N V V V W G	
CTTTTGCGCCGCTTTCATGATGCTGTGGAAAACAATCTCCGAATGTTGTTGTTGGGGA	8220
S G T P K R E F L H V D D M A S A S I Y	
AGTGGTACTCCAAAGCGTGAATTCTTACATGTAGATGATATGGCTTCTGCAAGCATTAT	8280
V M E M P Y D I W Q K N T K V M L S H I	
GTCATGGAGATGCCATACGATATATGGCAAAAAATACTAAAGTAATGTTGTCTCATATC	8340
N I G T G I D C T I C E L A E T I A K V	
AATATTGGAAACAGGTATTGACTGCACGATTTGTGAGCTTGCGGAAACAATAGCAAAAGTT	8400
V G Y K G H I T F D T T K P D G A P R K	
GTAGGTTATAAAGGCATATTACGTTTCGATACAACAAAGCCCGATGGAGCCCTCGAAAA	8460
L L D V T L L H Q L G W N H K I T L H K	
CTACTTGATGTAACGCTTCTTCATCAACTAGGTTGGAATCATAAAATTACCTTCACAAG	8520

Figure 6/7

End of orf8

G L E N T Y N W F L E N Q L Q Y R G *
GGTCTTGAAAATACATACAACCTGGTTTCTTGAAAACCAACTTCAATATCGGGGG TAATAA 8580

Start of orf9

M F L H S Q D F A T I V R S T P L I S I
TGTTTTACATTCCCAAGACTTTGCCACAATTGTAAGGTCTACTCCTCTATTTCTATAG 8640

D L I V E N E F G E I L L G K R I N R P
ATTTGATTGTGGAACGAGTTTGGCGAAATTTTGCTAGGAAAACGAATCAACCGCCCGG 8700

A Q G Y W F V P G G R V L K D E K L Q T
CACAGGCTATTGGTTCGTTCCGTGGTAGGGTGTGAAAGATGAAAAATTGCAGACAG 8760

A F E R L T E I E L G I R L P L S V G K
CCTTTGAACGATTGACAGAAATTGAACTAGGAATTCGTTTGCCTCTCTCTGTGGGTAAGT 8820

F Y G I W Q H F Y E D N S M G G D F S T
TTTATGGTATCTGGCAGCACTTCTACGAAGACAATAGTATGGGGGGAGACTTTTCAACGC 8880

H Y I V I A F L L K L Q P N I L K L P K
ATTATATAGTTATAGCATTCTTCTTAAATTACAACCAACATTTTGAAATTACCGAAGT 8940

S Q H N A Y C W L S R A K L I N D D D V
CACAACATAATGCTTATTGCTGGCTATCGCGAGCAAAGCTGATAAATGATGACGATGTGC 9000

H Y N C R A Y F N N K T N D A I G L D N
ATTATAATTGTGCGCATATTTTAAACAATAAAACAAATGATGCGATTGGCTTAGATAATA 9060

Start of orf10 End of orf9

M S D A P I I A V V M A G G T G S
K D I I C L M R Q *
AGGATATAATATGCTCTGATGCGCCAA TAATTGCTGTAGTTATGGCCGGTGGTACAGGCAG 9120

R L W P L S R E L Y P K Q F L Q L S G D
TCGTCTTTGGCCACTTTCTCGTGAACATATCCAAAGCAGTTTTTACAACCTCTCTGGTGA 9180

N T L L Q T T L L R L S G L S C Q K P L
TAACACCTTGTTACAAACGACTTTGCTACGACTTTCAGGCCTATCATGTCAAAAACCATT 9240

V I T N E Q H R F V V A E Q L R E I N K
AGTGATAACAAATGAACAGCATCGCTTTGTTGTGGCTGAACAGTTAAGGGAAATAAATAA 9300

L N G N I I L E P C G R N T A P A I A I
ATTAATGGTAATATATTCTAGAACCATGCGGGCGAAATACTGCACCAGCAATAGCGAT 9360

S A F H A L K R N P Q E D P L L L V L A
ATCTGCGTTTCATGCGTTAAAACGTAATCCTCAGGAAGATCCATTGCTTCTAGTTCTTGC 9420

A D H V I A K E S V F C D A I K N A T P
GGCAGACCACGTTATAGCTAAAGAAAGTGTTTTCTGTGATGCTATTAAAAATGCAACTCC 9480

I A N Q G K I V T F G I I P E Y A E T G
CATCGCTAATCAAGGTAAAATTGTAACGTTTGAATTATACCAGAATATGCTGAAACTGG 9540

Y G Y I E R G E L S V P L Q G H E N T G
TTATGGGTATATTGAGAGAGGTGAACATCTGTACCGCTTCAAGGGCATGAAAATACTGG 9600

F Y Y V N K F V E K P N R E T A E L Y M
TTTTTATTATGTAAATAAGTTTGTGCGAAAAGCCTAATCGTGAAACCGCAGAATTGTATAT 9660

T S G N H Y W N S G I F M F K A S V Y L
GACTTCTGGTAATCACTATTGGAATAGTGAATATTTCATGTTTAAGGCATCTGTTTATCT 9720

Figure 6/8

E E L R K F R P D I Y N V C E Q V A S S TGAGGAATTGAGAAAATTTAGACCTGACATTTACAATGTTTGTGAACAGGTTGCCTCATC	9780
S Y I D L D F I R L S K E Q F Q D C P A CTCATACATTGATCTAGATTTTATTCGATTATCAAAAGAACAATTTCAAGATTGTCCTGC	9840
E S I D F A V M E K T E K C V V C P V D TGAATCTATTGATTTTGCTGTAATGGAAAAACAGAAAAATGTGTTGTATGCCCTGTTGA	9900
I G W S D V G S W Q S L W D I S L K S K TATTGGTTGGAGTGACGTTGGATCTTGGCAATCGTTATGGGACATTAGTCTAAAATCGAA	9960
T G D V C K G D I L T Y D T K N N Y I Y AACAGGAGATGTATGTAAAGGTGATATATTAACCTATGATACTAAGAATAATTATATCTA	10020
S E S A L V A A I G I E D M V I V Q T K CTCTGAGTCAGCGTTGGTAGCCGCCATTGGAATTGAAGATATGGTTATCGTGCAAATAA	10080
D A V L V S K K S D V Q H V K K I V E M AGATGCCGTTCTTGTGTCTAAAAAGAGTGATGTACAGCATGTAAAAAAAATAGTCGAAAT	10140
L K L Q Q R T E Y I S H R E V F R P W G GCTTAAATTGCAGCAACGTACAGAGTATATTAGTCATCGTGAAGTTTCCGACCATGGGG	10200
K F D S I D Q G E R Y K V K K I I V K P AAAATTTGATTTCGATTGACCAAGGTGAGCGATACAAAGTCAAGAAAAATTATTGTGAAACC	10260
G E G L S L R M H H H R S E H W I V L S TGGTGAGGGGCTTTCTTTAAGGATGCATCACCATCGTTCTGAACATTGGATCGTGCTTTC	10320
G T A K V T L G D K T K L V T A N E S I TGGTACAGCAAAAGTAACCCCTTGGCGATAAACTAAACTAGTCACCGCAAATGAATCGAT	10380
Y I P L G A A Y S L E N P G I I P L N L ATACATTTCCCTTGGCGCAGCGTATAGTCTTGAGAATCCGGGCATAATCCCTCTTAATCT	10440
I E V S S G D Y L G E D D I I R Q K E R TATTGAAGTCAGTTCAGGGGATTATTGAGGAGGATGATATTATAAGACAGAAAGAAGC	10500
End of orf10 Start of orf11	
Y K H E D * M K S L T C F K A Y D I R TTACAAACATGAAGATTAACATATGAAATCTTTAACCTGCTTTAAAGCCTATGATAATTCG	10560
G K L G E E L N E D I A W R I G R A Y G CGGGAAATTAGGCGAAGAACTGAATGAAGATATTGCCTGGCGCATTGGGCGTGCTATGG	10620
E F L K P K T I V L G G D V R L T S E A CGAATTTCTCAAACCGAAAACCATTTGTTTAGGCGGTGATGTCCGCCTCACCAGCGAAGC	10680
L K L A L A K G L Q D A G V D V L D I G GTTAAACTGGCGCTTGCGAAAGGTTTACAGGATGCGGGCGTCGATGTGCTGGATATCGG	10740
M S G T E E I Y F A T F H L G V D G G I TATGTCCGGCACCAGAGATCTATTTCCGCCACGTTCCATCTCGGAGTGGATGGCGGCAT	10800
E V T A S H N P M D Y N G M K L V R E G CGAAGTTACCGCCAGCCATAACCCGATGGATTACAACGGCATGAAGCTGGTGCGGAAGG	10860
A R P I S G D T G L R D V Q R L A E A N GGCTCGCCCGATCAGCGGTGATACCGGACTGCGCGATGTCCAGCGTCTGGCAGAAGCCAA	10920
D F P P V D E T K R G R Y Q Q I N L R D TGACTTCCCTCTGTGCGATGAAACCAAACGTGGTTCGCTATCAGCAAATCAATCTGCGTGA	10980

Figure 6/9

A Y V D H L F G Y I N V K N L T P L K L CGCTTACGTTGATCACCTGTTCCGGTTATATCAACGTCAAAAACCTCACGCCGCTCAAGCT	11040
V I N S G N G A A G P V V D A I E A R F GGTGATCAACTCCGGAACGGCGCAGCGGTCCGGTGGTGGACGCCATTGAAGCCCGATT	11100
K A L G A P V E L I K V H N T P D G N F TAAAGCCCTCGGCGCACCGGTGAATTAATCAAAGTACACAACACGCCGACGGCAATTT	11160
P N G I P N P L L P E C R D D T R N A V CCCCAACGGTATTCCCTAACCCGCTGCTGCCGGAATGCCGCGACGACACCCGTAATGCGGT	11220
I K H G A D M G I A F D G D F D R C F L CATCAACACGGCGCGGATATGGGCATTGCCCTTTGATGGCGATTTTGACCGCTGTTTCCT	11280
F D E K G Q F I E G Y Y I V G L L A E A GTTTGACGAAAAAGGGCAGTTTATCGAGGGCTACTACATTGTTCGGCCTGCTGGCAGAAGC	11340
F L E K N P G A K I I H D P R L S W N T GTTCTCGAAAAAATCCCGGCGCAAGATCATCCACGATCCACGTCTCTCTCGAACAC	11400
V D V V T A A G G T P V M S K T G H A F CGTTGATGTGGTGACTGCCGCGAGCGGCACCCCGTAAATGTCGAAAACCGGACACGCCTT	11460
I K E R M R K E D A I Y G G E M S A H H TATTAAAGACGTATGCGCAAGGAAGACGCCATCTACGGTGGCGAAATGAGCGCTCACCA	11520
Y F R D F A Y C D S G M I P W L L V A E TTACTTCCGTGATTTTCGCTTACTGCGACAGCGCATGATCCCGTGGCTGCTGGTCGCCGA	11580
L V C L K G K T L G E M V R D R M A A F ACTGGTGTGCCTGAAAGGAAAAACGCTGGGCGAAATGGTGCAGCGACCGGATGGCGGCGTT	11640
P A S G E I N S K L A Q P V E A I N R V TCCGGCAAGCGGTGAGATCAACAGCAAACGGCGCAACCCGTTGAGGCAATTAATCGCGT	11700
E Q H F S R E A L A V D R T D G I S M T GGAACAGCATTTTAGCCGCGAGGCGCTGGCGGTGGATCGCACCGATGGCATCAGCATGAC	11760
F A D W R F N L R S S N T E P V V R L N CTTTGCCGACTGGCGCTTTAACCTGCGCTCCTCCAACACCGAACCAGGTGGTGGCGTTGAA	11820
V E S R G D V K L M E K K T K A L L K L TGTGGAATCACGCGTGATGTAAAGCTAATGAAAAGAAAAGCTAAAGCTCTTCTTAAATT	11880
End of orf11	
L S E *	
GCTAAGTGAGTGATTATTTACATTAAATCATTAAGCGTATTTAAGATTATATTAAAGTAAT	11940
GTTATTGCGGTATATGATGAATATGTGGGCTTTTTTATGTATAACGACTATACCGCAACT	12000
Start of H-repeat	
TTATCTAGGAAAAGATTAAATAGAAATAAAGTTTGTACTGACCAATTTGCATTTTCACGTC	12060
ACGATTGAGACGTTCCCTTTGCTTAAGACATTTTTTCATCGCTTATGTAATAACAAATGTG	12120
CCTTATATAAAAAGGAGAACAAAATGGAACTTAAAATAATTGAGACAATAGATTTTATT	12180
ATCCCTGTTTACGATATTATAGCCAAAGTTGTATCCTGCATCAGTCTGCAATATTTTAC	12240
GAGTGCTTTGTAACTGAATACATGTCTGCCATTTTCCAGATGATAACGACGTCATCGCA	12300
ATTGATGGTAAAACACTTCGGGCACACTTATGACAAGAGTCGTGCGAGAGGAGTGGTTCAT	12360

Figure 6/10

GTCATTAGTGCCTTTCAGCAATGCACAGTCTGGTCTCGGATAGATCAAGACGGATGAGA	12420
AACCTAATGCGTTCACAGTTATTCATGAACCTTCTAAAATGATGGGTATTAAAGGAAAAA	12480
TAATCATAACTGATGCGATGGCTTGCCAGAAAGATATTGCAGAGAAGATATAAAAAACAGA	12540
GATGTGATTATTTATTCGCTGTAAAAGGAAATAAGAGTCGGCTTAATAGAGTCTTTGAGG	12600
AGATATTTACGCTGAAAGAATTAAATAATCCAAAACATGACAGTTACGCAATTAGTGAAA	12660
AGAGGCACGGCAGAGACGATGTCCGTCTTCATATTGTTTGAGATGCTCCTGATGAGCTTA	12720
TTGATTTTCACGTTTGAATGGAAAGGGCTGCAGAATTTATGAATGGCAGTCCACTTTCTCT	12780
CAATAATAGCAGAGCAAAAAGAAAGATCCGAAATGACGATCAAAATATTATATTAGATCTG	12840
CTGCTTTTAACCGCAGAGAAGTTCCGCCACAGTAAATCGAAATCACTGGCGCATGGAGAATA	12900
AGTTGCACAGTAGCCTGATGTGGTAATGAATGAAATCGACTATAATATAAGAAGGCGAGT	12960
TGCATTGCAATGATTTTCTAGAATGCGGCACATCGCTATTAATATCTGACAATGATAATG	13020
TATTCAAGGCAGGATTATCATGTAAGATGCGAAAAGCAGTCATGGACAGAACTTCCTAG	13080
End of the H-repeat	
CGTCAGGCATTGCAGCGTCCGGGCTTTCATAATCTTGCAT TGGTTTGTGATAAGATATTTC	13140
Start of orf12	
TTTGGAGATGGGAAAATGAATTTGTATGGTATTTTGGTGTCTGGAAGTTATGGTAGAGAA	13200
T I P I L N Q Q I K Q E C G S D Y A L V	
ACAATACCCATCTCTAAATCAACAAATAAGCAAGAATGTGGTTCTGACTATGCTCTGGTT	13260
F V D D V L A G K K V N G F E V L S T N	
TTTGTGGATGATGTTTTTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGC'TTCAACCAAC	13320
C F L K A P Y L K K Y F N V A I A N D K	
TGCTTTCTAAAAGCCCCTTATTTAAAAAAGTATTTAATGTTGCTATTGCTAATGATAAG	13380
I R Q R V S E S I L L H G V E P I T I K	
ATACGACAGAGAGTGTCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAA	13440
H P N S V V Y D H T M I G S G A I I S P	
CATCCAAATAGCGTTGTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCC	13500
F V T I S T N T H I G R F F H A N I Y S	
TTTGTTACAATATCTACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCA	13560
Y V A H D C Q I G D Y V T F A P G A K C	
TACGTTGCACATGATTGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGT	13620
N G Y V V I E D N A Y I G S G A V I K Q	
AATGGATATGTTGTTATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAG	13680
G V P N R P L I I G A G A I I G M G A V	
GGTGTTCCTAATCGCCCACTTATTATGGCGCGGGAGCCATTATAGGTATGGGGGCTGTT	13740
V T K S V P A G I T V C G N P A R E M K	
GTCATAAAAGTGTTCCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAA	13800
End of orf12	
R S P T S I *	
AGATCGCCAACATCTATTTAATGGGAATGCGAAAACACGTTCCAAATGGGACTAATGTTT	13860

Figure 6/11

AAAATATATATAATTTCGCTAATTACTAAATTATGGCTTCTTTTAAAGCTATCCTTTAC 13920
TTAGTTATTACTGATACAGCATGAAATTTATAATACTCTGATACATTTTATACGTTATT 13980
CAAGCCGCATATCTAGCGGTAACCCCTGACAGGAGTAAACAATG 14024

ATGGCACAAAGTCATTAATACCAACAGCCTCTCGCTGATCACTCAAATAATATCAACAAG
 AACCAGTCTGCGCTGTGCGAGTTCTATCGAGCGTCTGTCTTCTGGCTTGCGTATTAACAGC
 GCGAAGGATGACGCCGCGGGTCAGGCGATTGCTAACCGTTTACTTCTAACATTAAAGGC
 CTGACTCAGGCTGCACGTAACGCCAACGACGGTATTTCTGTTGCACAGACCACTGAAGGC
 GCGCTGTCCGAAATCAACAACACTTACAGCGTATCCGTGAGCTGACGGTTCAGGCTTCT
 ACCGGGACTAACTCTGATTCCGATCTGGACTCCATTACAGGACGAAATCAAATCCCGTCTC
 GACGAAATTGACCGCGTATCCGGTCAGACCCAGTTCAACGGCGTGAACGTACTGGCAAAA
 GACGGTTCGATGAAAATTCAGGTAGGTGCGAACGACGGCCAGACTATCACTATTGATCTG
 AAGAAAATTGACTCTGATACGCTGGGGCTGAATGGTTTTAACGTGAATGGTTCCGGTACG
 ATAGCCAATAAAGCGGCGACCATTAGCGACCTGACAGCAGCGAAAATGGATGCTGCAACT
 AATACTATAACTACAACAAATAATGCGCTGACTGCATCAAAGGCCCTTGATCAACTGAAA
 GATGGTGACACTGTTACTATCAAAGCAGATGCAGCTCAAACGCCACGGTCTATACATAC
 AATGCATCTGCTGGTAACTTCTCATTAGTAATGTATCGAATAATACTTCAGCAAAAGCA
 GGTGATGTAGCAGCTAGCCTTCTCCCGCCGGCTGGGCAAACTGCTAGTGGTGTTCACAA
 GCAGCAAGCGGTGAAGTGAACCTTTGATGTTGATGCGAATGGTAAAATTACAATCGGAGGA
 CAGGAAGCCTATTTAACTAGTGATGGTAACTTAACTACAAACGATGCTGGTGGTGCGACT
 GCGGCTACGCTTGATGGTTTATTCAAGAAAGCTGGTGATGGTCAATCAATCGGGTTTAAT
 AAGACTGCATCAGTCACGATGGGGGAACAACCTTATAACTTTAAAACGGGTGCTGATGCT
 GGTGCTGCAACTGCTAACGCAGGGGTATCGTTCACTGATACAGCTAGCAAAGAAACCGTT
 TTAAATAAAGTGGCTACAGCTAAACAAGGCACAGCAGTTGCAGCTAACGGTGATACATCC
 GCAACAATTACCTATAAATCTGGCGTTCAGACGTATCAGGCGGTATTTGCCGCAGGTGAC
 GGTACTGCTAGCGCAAAATATGCCGATAATACTGACGTTTCTAATGCAACAGCAACATAC
 ACAGATGCTGATGGTGAAATGACTACAATTGGTTCATACACCACGAAGTATTCAATCGAT
 GCTAACAAACGGCAAGGTAACTGTTGATTCTGGAACCTGGTTCGGGTAAATATGCGCCGAAA
 GTCGGGGCTGAAGTATATGTTAGTGCTAATGGTACTTTAACAACAGATGCAACTAGCGAA
 GGCACAGTAACAAAAGATCCACTGAAAGCTCTGGATGAAGCTATCAGCTCCATCGACAAA
 TTCCGTTTCATCCCTGGGGGCTATCCAAAACCGTTTGGATTCCGCCGTCACCAACCTGAAC
 AACACCACTACCAACCTGTCTGAAGCGCAGTCCCGTATTACAGGACGCCGACTATGCGACC
 GAAGTGTCACATGTCGAAAGCGCAGATTATCCAGCAGGCCGGTAACTCCGTGCTGGCA
 AAAGCCAACCAGGTACCGCAGCAGGTTCTGTCTCTACTGCAGGGTTAA

Figure 7

AACAAATCTCAGTCTTCTCTTAGCTCTGCTATT
 GAGCGTCTGTCTTCTGGTCTGCGTATTAACAGCGCAAAAGACGATGCAGCAGGTCAGGCG
 ATTGCTAACCGTTTTACGGCAAATATTAAAGGTCTGACCCAGGCTTCCCGTAACGCAAAT
 GATGGTATTTCTGTTGCGCAGACCACTGAAGGTGCGCTGAATGAAATTAACAACAACCTG
 CAGCGTATTCGTGAACCTTCTGTTTCAAGCACTAACGGTACTAACTCTGACAGTGACCTG
 ACCTCCATCCAGTCCGAAATCCAGCAGCGTCTGAGTGAAATTGACCGTGTCTTCTGGTCAG
 ACTCAGTTTAACGGCGTTAAAGTGCTGGCTTCTGATCAGGATATGACTATTCAGGTTGGT
 GCAAACGACGGCGAAACAATTACTATTAAACTGCAGGAAATTAATTCCGACACACTGGGA
 TTATCTGGTTTTGGTATTAAAGATCCTACTAAATTTAAAGCCGCAACGGCTGAAACAAC
 TATTTTGGATCGACAGTTAAGCTTGCTGACGCTAATACACTTGATGCAGATATTACAGCT
 ACAGTTAAAGGCACTACGACTCCGGGCCAACGTGACGGTAATATTATGTCTGATGCTAAC
 GGTAAGTTGTACGTTAAAGTTGCCGGTTCAGATAAACCCGCTGAAAATGGTTATTATGAA
 GTTACTGTGGAGGATGATCCGACATCTCTGATGCAGGTAAGCTGAAGCTGGGGCTCTA
 GCGGGTACCCAGCCTCAAGCTGGTAATTTAAAGGAAGTCACAACGGTGAAAGGGAAGGGG
 GCTATTGATGTTTCAAGTTGGGTACTGATACCGCAACCGCTTCTATCACAGGTGCAAACTC
 TTTAAGTTAGAAGACGCCAATGGCAAAGATACTGGTTCATTTGCGTTGATTGGTGATGAC
 GGTAACAGTATGCAGCGAATGTTGATCAGAAAACAGGAGCAGTTTCCGTAAAAACAATG
 TCTTACACTGATGCTGACGGTGTCAAACACGACAATGTTAAAGTTGAACTGGGTGGAAGC
 GATGGCAAAACCGAAGTTGTAAGTCAACCGATGGCAAACTTACAGTGTTAGTGATTTA
 CAAGGTAAGAGCCTGAAAACCTGATTCTATTGCAGCAATTTCTACGCAGAAAACAGAAGAT
 CCTTTGGCTGCTATCGATAAAGCACTGTCTCAGGTTGACTCGTTGCGTTCTAACCTAGGT
 GCAATTCAAAATCGTTTCGACTCTGCCATCACCAACCTTGGCAACACCGTAAACAACCTG
 TCTTCTGCCCCGTAGCCGTATCGAAGATGCTGACTACGCGACCGAAGTGTCTAACATGTCT
 CGTGCGCAGATCCTGCAACAAGCGGGTACCTCTGTTCTGGCGCAG

Figure 8

AACAAATCTCAGTCTTCTCTGAGCTCCGCCATTGAACGTCTCTCTTC
 TGGCCTGCGTATTAACAGTGCTAAAGATGACGCAGCAGGTGAGGCGATTGCTAACCGTTT
 TACAGCAAATATTAAAGGTCTGACTCAGGCTTCCCGTAACGCGAATGATGGTATTTCTGT
 TGCGCAGACCCTGAAGGTGCGCTTTCTGAAATCAACAATAACTTACAGCGTATTCGTGA
 ATTGTCAGTACAGGCCACTAATGGTACAACTCTGACTCCGACCTGAATTCAATTCAGGA
 TGAAATTACACAACGCCTTAGTGAAATTGATCGTGTTCCTAACCAGACACAATTTAATGG
 TGTAAAAGTTCTGGCTTCTGATCAGACTATGAAAATTCAAGTAGGTGCGAACGATGGTGA
 AACCATTGAGATTGCCCTTGATAAAATTGATGCTAAAACCTTGGGGCTTGATAACTTTAG
 CGTAGCACCAGGAAAAGTTCCAATGTCCTCTGCGGTTGCACTTAAGAGCGAAGCCGCTCC
 TGACTTAAGGTAAATGCAACTGATGGTAGTGTGGGAGGTGCTAAAGCATTGGTAG
 CAATTATAAAAAATGCTGATGTTGAACTTATTTTGGTACCGGTAATGTACAAGATACAAA
 GGATACAACTGATGCGACCGGTACTGCAGGAACAAAAGTTTATCAAGTACAGGTGGAAGG
 GCAGACTTATTTTGGTGGTCAAGATAATAATACCAACACGAACGGTTTTACATTATTGAA
 AAAAACTCTACAGGTTATGAAAAAGTTCAGGTGGGTGGTAAGGATGTTTCAGTTAGCAA
 CTTTGGTGGTCTGTAACTGCATTTGTTGAAGATAATGGTTCTGCCACATCAGTTGATTT
 AGCTGCGGGTAAATGGGTAAAGCATTAGCTTATAATGATGCACCAATGTCTGTTATTT
 TGGGGGAAAAAACCTAGATGTCCACCAAGTACAAGATACCCAAGGGAATCCTGTACCTAA
 TTCATTTGCTGCTAAAACATCAGACGGCACCTACATTGCAGTAAATGTAGATGCCGCTAC
 AGGTAAACAGTCTGTTATTACTGATCCTAATGGTAAGGCAGTTGAATGGGCAGTAAAAAA
 TGATGGTTCTGCACAGGCAATTATGCGTGAAGATGATAAGGTTTATACAGCCAATATCAC
 GAATAAGACGGCAACCAAAGGTGCTGAACTCAGTGCCTCAGATTTGAAAGCCTTAGCAAC
 CACAAATCCATTATCCACATTAGACGAAGCTTTGGCAAAAGTTGATAAGTTGCGCAGTTC
 TTTGGGTGCAGTACAAAACCGTTTCGACTCTGCCATCACCAACCTTGGCAACACCGTAAA
 CAACCTGTCTTCTGCCCCGTAGCCGTATAGAAGATGCTGACTACGCAACCGAAGTGTCTAA
 CATGTCTCGTGCGCAGATCCTGCAACAAGCGGGTACCTCTGTTCTGGCACAG

Figure 9

AACAAAAACCAGTCTGCGCTGTGACTTCTATCGAG
CGCCTCTCTTCTGGTCTGCGTATTAAACAGCGCTAAAGATGACGCCGCGGGCCAGGCGATT
GCTAACCGCTTTACTTCTAACATCAAAGGTCTGACTCAGGCCGCACGTAACGCCAACGAC
GGTATTTCTCTGGCGCAGACGGCTGAAGGCGCGCTGTCAGAGATTAACAACAACCTTGCAG
CGTATTCGTGAACTGACCGTTCAGGCCTCTACCGGCACGAACTCTGATTCCGACCTGTCT
TCTATTCAGGACGAAATCAAATCCCGTCTTGATGAAATTGACCGTGTATCTGGTCAGACC
CAGTTCAACGGTGTGAACGTGCTGTCGAAAAACGATTTCGATGAAGATTCAGATTGGTGCC
AATGATAACCAGACGATCAGCATTGGCTTGCAACAAATCGACAGTACCACTTTGAATCTG
AAAGGATTTACCGTGTCCGGCATGGCGGATTTACGCGCGCGAAACTGACGGCTGCTGAT
GGTACAGCAATTGCTGCTGCGGATGTCAAGGATGCTGGGGGTAAACAAGTCAATTTACTG
TCTTACACTGACACCGCTCTAACAGTACTAAATATGCGGTGCTTGATTCTGCAACCGGT
AAATACATGGAAGCCACTGTAGTCATTACCGGTACGGCGCGCGGTAAGTGTGGTGCA
GCGGAAGTGGCGGGAGCCGCTACAGCCGATCCGTTAAAGCACTGGATGCCGCAATCGCT
AAAGTCGACAAATTCCGCTCCTCCCTCGGTGCCGTTCAAACCGTCTGGATTCTGCGGTC
ACCAACCTGAACAACACCACCACCAACCTGTCTGAAGCGCAGTCCCCTATTTCAGGACGCC
GACTATGCGACCGAAGTGTCCAACATGTGCGAAAGCGCAGATTATCCAGCAGGCGGGCAAC TCCGTGCTGTCTAA

Figure 10

AACAAAAACCAGTCTGCGCTGTCGACTTCTAT
CGAGCGCCTCTCTTCTGGTCTGCGTATTAACAGCGCTAAAGATGACGCCGCGGGCCAGGC
GATTGCTAACCGCTTCACTTCTAACATCAAAGGTCTGACTCAGGCCGCACGTAACGCCAA
CGACGGTATCTCTTGGCGCAGACCACTGAAGGCGCGCTGTCTGAAATCAACAACAACCTT
GCAGCGTGTGCGTGAGTTGACCGTTCAGGCGACGACCGGGACTAACTCTGATTCTGACCT
GTCTTCTATTAGGACGAAATCAAATCCCGTCTGGATGAAATTGATCGCGTTTCCGGTCA
GACCCAGTTCAACGCGGTGAATGTGCTGGCGAAAGATGGTTCGATGAAGATTCAGGTTGG
CGCGAATGATGGGCAGACTATTAGCATTGATTTGCAGAAGATTGACTCTTCTACATTAGG
ACTGAACGGTTTCTCCGTTTCGGGTCAGTCACTTAACGTTAGTGATTCCATTACTCAAAT
TACCGGTGCCGCCCGGGACAAAACCTGTTGGTGTGATTTCACTGCTGTTGCGAAAGATCT
GACTACTGCGACAGGTAAACAGTCGATGTTTCTAGCCTGACGTTACACAACACTCTGGA
TGCGAAAGGGGCTGCTACATCACAGTTCGTCGTTCAATCCGGCAATGATTTCTACTCCGC
GTCGATTAATCATAACAGACGGCAAAGTCACGTTGAATAAAGCCGATGTCGAATACACAGA
CACCGATAATGGACTAACGACTGCGGCTACTCAGAAAGATCAACTGATTAAAGTTGCCGC
TGACTCTGACGGCTCGGCTGCGGGATATGTAACATTCCAAGGTAAAACTACGCTACAAC
GGTTTCAACGGCACTTGATGATAATACTGCGGCAAAAGCAACAGATAATAAAGTTGTTGT
TGAATTATCAACAGCAAAACCGACTGCACAGTTCTCAGGGGCTTCTTCTGCTGATCCACT
GGCACTTTTAGACAAAGCTATTGCACAGGTTGATACTTCCGCTCCTCCCTCGGTGCGGT
GCAAAACCGTCTGGATTCCGCAGTAACCAACCTGAACAACACCACCACCAACCTGTCTGA
AGCGCAGTCCCGTATTAGGACGCCGACTATGCTACAGAAGTGTCACATGTGAAAGC
GCAGATCATCCAGCAGGCAGGTAACCTCGGTGCTGTCCAAA

Figure 11

ATGGCACAAGTCATTAATACCAACAGCCTCTCGC
TGATCACTCAAAATAATATCAACAAGAACCAGTCTGCGCTGTGAGTTCTATCGAGCGTC
TGTCTTCTGGCTTGCGTATTAAACAGCGCGAAGGATGACGCCGCGGTGAGGCGATTGCTA
ACCGTTTACTTCTAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAACGACGGTA
TTTCTGTTGCGCAGACCACCGAAGGCGCGTGTCCGAAATTAACAACAACCTTACAGCGTA
TTCGTGAAGTACCGGTTTACGGCTTCTACCGGGACTAACTCTGATTCCGGATCTGGACTCCA
TTCAGGACGAAATCAAATCCCGTCTCGACGAAATTGACCGCGTATCCGGTCAGACCCAGT
TCAACGGCGTGAACGTACTGGCAAAAGACGGTTCGATGAAAATTCAGGTTGGTGCGAATG
ACGCCCAGACTATCACTATTGATCTGAAGAAAATTGACTCTGATACGCTGGGGCTGAATG
GGTTTAAATGTGAACGGCAAAGGGGAAACGGCTAATACGGCAGCAACCCTGAAAGATATGT
CTGGATTACAGCTGCGGCGGCACCAGGGGAACTGTTGGTGTAACCTCAATATACTGACA
AATCGGCTGTAGCAAGTAGCGTAGATATTCTAAATGCTGTTGCTGGCGCAGATGGAAATA
AAGTTACAAGTAGCGCCGATGTTGGTTTTGGTACACCAGCCGCTGCTGTAACCTATACCT
ACAATAAGACACTAATTATATTCGCCGCTTCTGATGATATTTCCAGCGCTAACCTGG
CTGCTTTCCCTCAATCCTCAGGCCGGAGATACGACTAAAGCTACAGTTACAATTGGTGGCA
AAGATCAAGATGTAAACATCGATAAATCCGGTAATTTAACTGCTGCTGATGATGGCGCAG
TACTTTATATGGATGCTACCGGTAACCTAACTAAAAATAATGCTGGTGGTGATACACAAG
CTACTTTGGCTAAACTTGCTACTGCTACTGGTGCTAAAGCCGCGACCATCCAAACTGATA
AAGGAACATTACACAGTGACGGTACAGCGTTTGATGGTGATCAATGTCCATTGATACCA
ATACATTTGCAAATGCAGTAAAAAATGACACTTATACTGCCACTGTAGGTGCTAAGACTT
ATAGCGTAACAACAGGTTCTGCTGCTGCAGACACCGCTTATATGAGCAATGGGGTTCTCA
GTGATACTCCGCCAACTTACTATGCACAAGCTGATGGAAGTATCACAACTACTGAGGATG
CGGCTGCCGGTAACTGGTCTACAAAGGTTCCGATGGTAAGTTAACAACGGATACGACTA
GCAAAGCAGAATCAACATCAGATCCGCTGGCAGCTCTTGACGACGCTATCAGCCAGATCG
ACAAATTCCGCTCCTCCCTGGGTGCGGTGCAAAACCGTCTGGATTCCGCAGTGACCAACC
TGAACAACACCACTACCAACCTGTCTGAAGCGCAGTCCCGTATTGAGGACGCCGACTATG
CGACCGAAGTGTCACATGTGAAAGCGCAGATTATCCAGCAGGCCGGTAACTCCGTGC
TGGCAAAAGCTAACAGGTTCCGCGCAGGTTCTGTCTCTGCTGCAGGGTTAA

Figure 12

ATGGCACAAG TCATTAATAC CAACAGCCTC TCGCTGATCA CTCAAAATAA TATCAACAAG
 AACCAGTCTG CGCTGTCGAG TTCTATCGAG CGTCTGTCTT CTGGCTTGCG TATTAACAGC
 GCGAAGGATG ACGCCGCGGG TCAGGCGATT GCTAACCGTT TTAATTCTAA CATTAAAGGC
 CTGACTCAGG CTGCACGTAA CGCCAACGAC GGTATTTCTG TTGCACAGAC CACCGAAGGC
 GCGCTGTCTG AAATCAACAA CAACTTACAG CGTATCCGTG AGCTGACGGT TCAGGCTTCT
 ACCGGAAC TA CTCTGATTG GGATCTGGAC TCCATTGAGG ACGAAATCAA ATCCCGTCTT
 GATGAAATTG ACCGCGTATC CGGCCAGACC CAGTTCAACG GCGTGAACGT ACTGGCAAAA
 GACGGTTCGA TGAAAATTCA GGTGTTGTCG AATGACGGTG AACTATCAC TATCGACCTG
 AAGAAAATCG ATTCTGATAC TCTGGGTCTG AATGGTTTTA ACGTAAATGG TAAAGGTACT
 ATTACCAACA AAGCTGCAAC GGTAAGTGAT TTAATTCTG CTGGCGCGAA GTTAAACAC
 CACGACAGGT CTTTATGATC TGAAAACCGA AAATACCTTG TTAATTACCG ATGCTGCATT
 CGATAAATTA GGAATGGCG ATAAAGTCAC CGTTGGCGGC GTAGATTATA CTTACAACGC
 TAAATCTGGT GATTTTACTA CCACCAAATC TACTGCTGGT ACGGGTGTAG ACGCCGCGGC
 GCAGGCTACT GATTCAGCTA AAAAACGTGA TGCCTTAGCT GCCACCTTC ATGCTGATGT
 GGGTAAATCT GTTAATGGTT CTTACACCAC AAAAGATGGT ACTGTTTCTT TCGAAACGGA
 TTCAGCAGGT AATATCACC A TCGGTGGAAG CCAGGCATAC GTAGACGATG CAGGCAACTT
 GACGACTAAC AACGCTGGTA GCGCAGCTAA AGCTGATATG AAAGCGCTGC TTAAAGCCGC
 GAGCGAAGGT AGTGACGGTG CTTCTCTGAC ATTCAATGGC ACTGAATATA CTATCGCAAA
 AGCAACTCCT GCGACAACCT CTCCAGTAGC TCCGTTAATC CCTGGTGGGA TTACTTATCA
 GGCTACAGTG AGTAAAGATG TAGTATTGAG CGAAACCAA GCGGCTGCCG CGACATCTTC
 AATTACCTTT AATTCCGGTG TACTGAGCAA AACTATTGGG TTTACCGCGG GTGAATCCAG
 TGATGCTGCG AAGTCTTATG TGGATGATAA AGGTGGTATT ACTAACGTTG CCGACTATAC
 AGTCTCTTAC AGCGTTAACA AGGATAACGG CTCTGTGACT GTTGCCGGGT ATGCTTCAGC
 GACTGATACC AATAAAGATT ATGCTCCAGC AATTGGTACT GCTGTAAATG TGAATCCGC
 GGGTAAAATC ACTACTGAGA CTACCAGTGC TGTTCTGCA ACGACCAACC CGCTTGCTGC
 CCTGGACGAC GCTATCAGCT CCATCGACAA ATTCCGTTCT TCCCTGGGTG CTATCCAGAA
 CCGTCTGGAT TCCGAGTCA CCAACCTGAA CAACACCACT ACCAACCTGT CTGAAGCGCA
 GTCCCGTATT CAGGACGCCG ACTATGCGAC CGAAGTGTC AACATGTCGA AAGCGCAGAT
 TATCCAGCAG GCCGGTAACT CCGTGCTGGC AAAAGCCAAC CAGGTACCGC AGCAGGTTCT
 GTCTCTGCTG CAGGGTTAA

Figure 13

AACAAATCTCAGTCTTCTCTTAGCTCTGCTA
 TTGAGCGTCTGTCTTCTGGTCTGCGTATTAACAGCGCAAAAGACGATGCAGCAGGTCAGG
 CGATTGCTAACCGTTTTACGGCAAATATTAAAGGTCTGACCCAGGCTTCCCGTAACGCAA
 ATGATGGTATTTCTGTTGCGCAGACCCTGAAGGTGCGCTGAATGAAATTAACAACAACC
 TGCAGCGTATTCGTGAACCTTCTGTTTCAGGCAACTAACGGTACTAACTCTGACAGCGATC
 TTTCTTCTATCCAGGCTGAAATTACTCAACGTCTGGAAGAAATTGACCGTGTATCTGAGC
 AAACCTCAGTTTAACGGCGTGAAAGTCCTTGCTGAAAATAATGAAATGAAATTCAGGTTG
 GTGCTAATGATGGTGAAACCATCACTATCAATCTGGCAAAAATTGATGCGAAAACCTCTCG
 GCCTGGACGGTTTTAATATCGATGGCGCGCAGAAAGCAACAGGCAGTGACCTGATTTCTA
 AATTTAAAGCGACAGGTACTGATAATTATGATGTTGGCGGTAAAACTTATACCGTGAATG
 TGGAGAGCGGCGCGGTTAAGAATGATGCTAATAAAGATGTTTTTGTAAGCGCAGCTGATG
 GATCGCTGACGACCAGTAGTGATACTAAAGTATCCGGTGAAAGTATTGATGCAACAGAAC
 TAGCGAAACCTTGCAATAAAATTAGCTGACAAAGGCTCCATTGAATACAAGGGCATTACAT
 TTAATAACAACACTGGCGCAGAGCTTGATGCTAATGGTAAAGGTGTTTTGACCGCAAATA
 TTGATGGTCAAGATGTTCAATTTACTATTGACAGTAATGCACCCACGGGTGCCGGCGCAA
 CAATAACTACAGACACAGCTGTTTACAAAAACAGTGCGGGCCAGTTCACCACTACAAAAG
 TGGAAAAATAAGCGCAACACTCTCTGATCTGGATCTTAATGCAGCCAAGAAAACAGGTA
 GCACTTTAGTTGTAAATGGCGCCACCTACAATGTCAGCGCAGATGGTAAAACGGTAACTG
 ATACTACTCCTGGTGCCCCCTAAAGTGATGTATCTGAGCAAATCAGAAGGTGGTAGCCCGA
 TTCTGGTAAACGAAGATGCAGCAAAATCGTTGCAATCTACCACCAACCCGCTCGAAACTA
 TCGACAAGGCATTGGCTAAAGTTGACAATCTGCGTTCTGACCTCGGTGCAGTACAAAACC
 GTTTCGACTCTGCCATCACCACCTTGCGAACACCGTAAACAACCTGTCTTCTGCCCGTA
 GCCGTATCGAAGATGCTGACTACGCGACCGAAGTGTCTAACATGTCTCGTGCGCAGATCC
 TGCAACAAGCGGGTACCTCTGTTCTGGCGCAG

Figure 14

ATGGCACAAGTCATTAATACCAACAGCCTCTCG
 CTGATCACTCAAAATAATATCAACAAGAACCAGTCTGCGCTGTCGAGTTCTATCGAGCGT
 CTGTCTTCTGGCTTGCGTATTAACAGCGCGAAGGATGACGCCGCGGTCAGGCGATTGCT
 AACCGTTTTACTTCTAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAACGACGGT
 ATTTCCGTTGCACAGACCACTGAAGGCGCGCTGTCCGAAATTAACAACAACCTTACAGCGT
 ATTCGTGAACTGACGGTTTACGGCTTCTACCGGGACTAACTCCGATTCCGATCTGGACTCC
 ATTCAGGACGAAATCAAATCCCGTCTGGACGAAATGACCGCGTATCCGGCCAGACCCAG
 TTCAACGGCGTGAACGTGCTGTCCAAAGATGGCTCGATGAAAATTAGGTTCGGCGCGAAC
 GATGGCGAAACGATTACTATTGATCTGAAGAAAATTGACTCTGATACGCTGAATCTGGCT
 GGTTTTAACGTTAACGGTAAAGGTTCTGTAGCGAATACAGCTGCGACAAGCGACGATTTA
 AAACCTGGCTGGTTTCACTAAGGGCACCACAGATACCAATGGCGTGACCGCGTATACAAAC
 ACAATTAGTAATGACAAAGCCAAAGCTTCCGATCTGTTAGCTAATATCACCGATGGATCA
 GTGATCACTGGGGGAGGGGCAAACGCTTTTGGCGTGGCTGCAAAGAATGGTTACACCTAT
 GATGCAGCAAGTAAATCTTATAGTTTTGCTGCAGATGGTGCCGATTACGCGAAGACGTTA
 AGCATCATTAATCCAAACACCGGTGATTCGTGCGAGGCGACAGTGACTATTGGTGGTAAA
 GAGCAGAAAAGTTAATATTTCCAGGATGGAAAAATTACTGCGGCAGATGATAATGCGACG
 CTGTATTTAGATAAAACAGGGAAAATTGACAAAAACGAATGCAGGTAACGATACCGCAGCG
 ACTTGGGATGGTTTAAATTTCAACAGCGATTCTACCGGTGCGGTTCCAGTTGGGGTTGCA
 ACTACAATTACAATTACTTCTGGTACAGCTTCCGGAATGTCTGTTAGTCCGCGAGGAGCA
 GGAATTCAGACCTCAACAAATTTCTAGATTCTTGCAAGGTGGTGCATTTGCGGCTAAGGTA
 AGTATTGAGGGAGGCGCTGTACAGACATTTTGGTAGCAAGTAATGGAAACATAACAGCG
 GCTGATGGTAGTGCACTTTATCTTGATGCGACTACTGGTGGATTCACTACAACGGCTGGA
 GGAAATACAGCTGCTTCGTTAGATAATTTAATTGCTAACAGTAAGGATGCTACCTTAACC
 GTAACCTCAGGTACCGGCCAGAACACTGTTTATAGCACAAACAGGAAGTGGCGCTCAGTTC
 ACCAGTTTAGCAAAAGTAGACACAGTCAATGTCACCAACGCACATGTAGTGCCGAAGGT
 ATGGCAAATCTGACAAAAAGCAATTTTACCATTGATATGGGCGGTACAGGTACAGTAACT
 TACACAGTTTCCAATGGGGATGTGAAAGCTGCTGCAAATGCTGATGTTTATGTGCAAGAT
 GGTGCACTTTCAGCCAATGCTACAAAAGATGTAACCTACTTTGAACAAAAAATGGGGCT
 ATTACCAACAGCACCGGTGGTACCATCTATGAAACAGCTGATGGTAAGTTAACAACAGAA
 GCTACTACTGCATCCAGTTCCACCGCCGATCCCCTGAAAGCTCTGGACGAAGCCATCAGC
 TCCATCGACAAATTCCGCTCCTCCCTCGGTGCGGTGCAAAACCGTCTGGATTCCGCGGTC
 ACCAACCTGAACAACCACTACCAACCTGTCCGAAGCGCAGTCCCGTATTAGGACGCC
 GACTATGCGACCGAAGTGTCCAACATGTGAAAGCGCAGATCATCCAGCAGGCCGGTAAC
 TCCGTGCTGGCAAAAGCTAACAGGTACCGCAGCAGGTTCTGTCTCTGCTGCAGGGTTAA

Figure 15

ATGGCACAAAGTCATTAATACCAACAGC
CTCTCGCTGATCACTCAAAATAATATCAACAAGAACCAGTCTGCGCTGTCGAGTTCTATC
GAGCGTCTGTCTTCTGGCTTGCGTATTAACAGCGCGAAGGATGACGCCGCGGGTCAGGCG
ATTGCTAACCGTTTTACTTCTAACATTAAAGGCCCTGACTCAGGCTGCACGTAACGCCAAC
GACGGTATTTCTGTTGCGCAGACCACCGAAGGCGCGCTGTCCGAAATTAACAACAACCTTA
CAGCGTGTGCGTGAGCTGACTGTTTCAGGCGACCACCGGTACTAACTCTGAGTCTGACCTG
TCTTCTATCCAGGACGAAATCAAATCTCGCCTGGAAGAGATTGATCGTGTTCAGTCAG
ACTCAATTTAACGGCGTGAATGTTTTGGCTAAAGATGGGAAAATGAACATTCAGGTTGGG
GCAAAATGATGGACAGACTATCACTATTGATCTGAAAAAGATCGATTCTCTACACTAAAC
CTCTCCAGTTTTGATGCTACAAACTTGGGCACCAGTGTTAAAGATGGGGCCACCATCAAT
AAGCAAGTGGCAGTAGGTGCTGGCGACTTTAAAGATAAAGCTTCAGGATCGTTAGGTACC
CTAAAATTAGTTGAGAAAGACGGTAAGTACTATGTAAATGACACTAAAAGTAGTAAGTAC
TACGATGCCGAAGTAGATACTAGTAAGGGTAAATTAACCTCAACTCTACAAATGAAAGT
GGAACTACTCCTACTGCAGCGACGGAAGTAACTACTGTTGGCCGCGATGTAAAATTGGAT
GCTTCTGCACTTAAAGCCAACCAATCGCTTGTCGTGTATAAAGATAAAAGCGGCAATGAT
GCTTATATCATTCAGACCAAAGATGTAACAATAATCAATCAACTTTCAATGCCGCTAAT
ATCAGTGATGCTGGTGTCTTATCTATTGGTGCATCTACAACCGCGCCAAGCAATTTAACA
GCTAACCCGCTTAAAGGCTCTTGATGATGCAATTGCATCTGTTGATAAATCCGCTCTTCT
CTCGGTGCCGTTTCAGAACCGTCTGGATTCTGCCATTGCCAACCTGAACAACACCACTACC
AACCTGTCTGAAGCGCAGTCCCGTATTCAGGACGCTGACTATGCGACCGAAGTGTCCAAC
ATGTGCAAAGCGCAGATTATCCAGCAGGCCGGTAACTCCGTGCTGGCAAAAGCCAACCAG
GTACCGCAGCAGGTTCTGTCTCTGCTGCAGGGTTAA

Figure 16

AACAAATCTCAGTCTTCTCTGAGCTCCGCCAT
TGAACGTCTCTCTTCTGGCTGCGTATTAAAGGTGCTAAAGATGACGCAGCAGGTCAGGC
GATTGCTAACCGTTTTACAGCAAATATTAAAGGTCTGACTCAGGCTTCCCGTAACGCGAA
TGATGGTATTTCTGTTGCGCAGACCACTGAAGGTGCGCTGAATGAAATTAACAACAACCT
GCAGCGTGACGTGAACTGACTGTTCAAGGCAACTAACGGTACTAACTCTGACAGCGATCT
TTCTTCTATCCAGGCTGAAATTACTCAACGTCTGGAAGAAATTGACCGTGATCTGAGCA
AACTCAGTTTAAACGGCGTGAAAGTCCTTGCTGAAAATAATGAAATGAAAATTCAGGTTGG
TGCTAATGATGGTGAAACCATCACTATCAATCTGGCAAAAATTGATGCGAAAACCTCTCGG
CCTGGACGGTTTTAATATCGATGGCGCGCAGAAAGCAACTGGCAGTGACCTGATTCTTAA
ATTTAAAGCGACAGGTACTGATAACTATGATGTTGGCGGTGATGCTTATACTGTTAACGT
AGATAGCGGAGCTGTTAAAGATACTACAGGGAATGATATTTTGTAGTGCAGCAGATGG
TTCCTGACAACCTAAATCTGACACAAACATAGCTGGTACAGGGATTGATGCTACAGCACT
CGCAGCAGCGGCTAAGAATAAAGCACAGAATGATAAATTCACGTTTAAATGGAGTTGAATT
CACAACAACAACCTGCAGCGGATGGCAATGGGAATGGTGATATTTCTGCAGAAATTGATGG
TAAGTCAGTGACATTTACTGTGACAGATGCTGACAAAAAGCTTCTTTGATTACGAGTGA
GACAGTTTACAAAAATAGCGCTGGCCTTTATACGACAACCAAAGTTGATAACAAGGCTGC
CACACTTCCGATCTTGATCTCAATGCAGCTAAGAAAAACAGGAAGCACGTTAGTTGTTAA
CGGTGCAACTTACGATGTTAGTGCAGATGGTAAACGATAACGGAGACTGCTTCTGGTAA
CAATAAAGTCATGTATCTGAGCAAATCAGAAGGTGGTAGCCCGATTCTGGTAAACGAAGA
TGCAGCAAAATCGTTGCAATCTACCACCAACCCGCTCGAAACTATCGACAAAGCATTGGC
TAAAGTTGACAATCTGCGTTCTGACCTCGGTGCAGTACAAAACCGTTTCGACTCTGCTAT
CACCAACCTTGGCAACACCGTAAACAACCTGTCTTCTGCCCGTAGCCGTATCGAAGATGC
TGACTACGCGACCGAAGTGTCTAACATGTCTCGTGCGCAGATCTGCAACAAGCGGGTAC
CTCTGTTCTGGCGCAG

Figure 17

ATGGCACAAGTCATTAATACCAACAGCCTCTCGCTGATCA
 CTCAAATAATATCAACAAGAACAGTCTGCGCTGTCGAGTTCTATCGAGCGTCTGTCTT
 CTGGCTTGCGTATTAACAGCGCGAAGGATGACGCAGCGGGTCAGGCGATTGCTAACCGTT
 TCACCTCTAACATTAAGGCCCTGACTCAGGCGGCCCGTAACGCCAACGACGGTATCTCCG
 TTGCGCAGACCACCGAAGGCGCGCTGTCCGAAATCAACAACAATTACAGCGTATCCGTG
 AACTGACGGTTTCAAGCTTCTACCGGGACTAACTCCGATTCCGATCTGGACTCCATTCAAG
 ACGAAATCAAATCCCGTCTGGACGAAATTGACCGCGTATCTGGCCAGACCCAGTTCAACG
 GCGTGAACGTACTGGCGAAAGACGGTTCAATGAAAATTAGGTTGGTGCGAATGACGGCC
 AGACTATCACGATTGATCTGAAGAAAATTGACTCAGATACGCTGGGGCTGAATGGTTTTA
 ACGTGAATGGTTCCGGTACGATAGCCAATAAAGCGGCGACCATTAGCGACCTGACAGCAG
 CGAAAATGGATGTGCAACTAATACTATACTACAACAAATAATGCGCTGACTGCATCAA
 AGGCGCTTGATCAACTGAAAGATGGTGACACTGTTACTATCAAAGCAGATGCTGCTCAA
 CTGCCACGGTTTATACATACAATGCATCAGCTGGTAACTTCTCATTAGTAATGTATCGA
 ATAATACTTCAGCAAAAGCAGGTGATGTAGCAGCTAGCCTTCTCCCGCCGGCTGGGCAA
 CTGCTAGTGGTGTATTATAAAGCAGCAAGCGGTGAAGTGAACCTTTGATGTTGATGCGAATG
 GTAAAATCACAATCGGAGGACAGAAAGCATATTTAACTAGTGATGGTAACTTAACTACAA
 ACGATGCTGGTGGTGCGACTGCGGCTACGCTTGATGGTTTATTCAAGAAAGCTGGTGATG
 GTCAATCAATCGGGTTTAAGAAGACTGCATCAGTCACGATGGGGGAACAACCTTATAACT
 TTAACACGGGTGCTGATGCTGATGCTGCAACTGCTAACGCAGGGGTATCGTTCACTGATA
 CAGCTAGCAAAGAAACCGTTTTTAAATAAAGTGGCTACAGCTAAACAAGGCAAAGCAGTTG
 CAGCTGACGGTGATACATCCGCAACAATTACCTATAAATCTGGCGTTCAGACGTATCAGG
 CTGTATTTGCCGAGGTGACGGTACTGCTAGCGCAAAATATGCCGATAAAGCTGACGTTT
 CTAATGCAACAGCAACATACACTGATGCTGATGGTGAAATGACTACAATTGGTTCATACA
 CCACGAAGTATTCAATCGATGCTAACAACGGCAAGGTAACCTGTTGATTCTGGAACCTGGTA
 CGGGTAAATATGCGCCGAAAGTAGGGGCTGAAGTATATGTTAGTGCTAATGGTACTTTAA
 CAACAGATGCAACTAGCGAAGGCACAGTAACAAAAGATCCACTGAAAGCTCTGGATGAAG
 CTATCAGCTCCATCGACAAATTCCGTTCTTCCCTGGGTGCTATCCAGAACCGTCTGGATT
 CCGCAGTCACCAACCTGAACAACACCACTACCAACCTGTCCGAAGCGCAGTCCCGTATTC
 AGGACGCCGACTATGCGACCGAAGTGTCCAACATGTGAAAGCGCAGATCATTAGCAGG
 CCGGTAACCTCCGTGCTGGCAAAAGCCAACAGGTACCGCAGCAGGTTCTGTCTCTGCTGC AGGGTTAA

Figure 18

ATGGCACAAAGTCATTAATACCAACAGCCTCTCGCTGATCACTCAAAATA
 ATATCAACAAGAACCAGTCTGCGCTGTCGAGTTCTATCGAGCGTCTGTCTTCTGGCTTGC
 GTATTAACAGCGCGAAGGATGACGCCGAGGTGAGGCGATTGCTAACCGTTTTACTTCTA
 ACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAACGACGGTATTTCCGTTGCGCAGA
 CCACTGAAGGTGCGCTGTCCGAAATCAACAACAACCTTACAGCGTATTCGTGAGCTGACGG
 TTCAGGCTTCTACCGGGACTAACTCCGATTCTGACCTGGACTCCATCCAGGACGAAATCA
 AGTCTCGTCTGGACGAAATTGACCGCGTATCCGGTCAGACCCAGTTCAACGGCGTGAACG
 TGCTGGCGAAAGACGGTTCGATGAAAATTGAGTTGGTGCGAATGACGGCCAGACTATCA
 CGATTGATCTGAAGAAAATTGACTCAGATACGCTGGGGCTGAGTGGGTTTAATGTGAATG
 GTGGCGGGGCTGTTGCTAACACTGCTGCATCTAAAGCTGACTTGGTAGCTGCTAATGCAA
 CTGTGGTAGGCAACAAATATACTGTGAGTGCGGGTTACGATGCTGCTAAAGCGTCTGATT
 TGCTGGCTGGAGTTAGTGATGGTGATACTGTTTACGGCAACCATTAATAACGGCTTCGGAA
 CGGCGGCTAGTGCAACGAATTACAAGTATGACAGTGCAAGTAAGTCTTACTCTTTTGATA
 CCACAACGGCTTCAGCTGCCGATGTTGAGAAATATTTGACCCCGGGCGTTGGTGATACCG
 CTAAGGGCACTATTACTATCGATGGTTCGACAGGATGTTTACAGTACAGCAGTGATGGTA
 AAATTACGTCAAGCAATGGAGATAAACTTTACATTGATACAACTGGGCGCTTAACGAAAA
 ACGGCTTTAGTGCTTCTTTGACTGAGGCTAGTCTGTCCACACTTGACGCCAATAATACCA
 AAGCGACAACCATTGACATTGGCGGTACCTCTATCTCCTTTACCGGTAATAGTACTACGC
 CGAACACTATTACTTATTTCAGTAACAGGTGCAAAAGTTGATCAGGCAGCTTTCGATAAAG
 CTGTATCAACCTCTGGAAACGATGTTGATTTCACTACCGCAGGTTATAGCGTCGACGGCG
 CAACTGGCGCTGTAAACAAAAGGTGTTGCTCCGGTTTATATTGATAACAACGGGGCGTTGA
 CCACATCTGATACTGTAGATTTTTATCTACAGGATGATGGTTCAGTGACTAACGGCAGCG
 GTAAGGCAGTTTATAAAGATGCTGACGGTAAATTGACGACAGATGCTGAAACTAAAGCTG
 CAACCACCGCCGATCCCCTGAAAGCTCTGGACGAAGCCATCAGCTCCATCGACAAATTCC
 GCTCCTCCCTCGGTGCGGTGCAGAACCGTCTGGATTCCGCGGTACCAACCTGAACAACA
 CCACTACCAACCTGTCTGAAGCGCAGTCCCGTATTGAGGACGCTGACTATGCGACCGAAG
 TATCCAACATGTGAAAGCGCAGATCATCCAGCAGGCGGTAACCTCCGTGCTGGCAAAAG
 CTAACCAGGTACCACAGCAGGTTCTGTCTCTGCTGCAGGGTTAA

Figure 19

ATGGCACAAGTCATTAATACCAACAGC
 CTCTCGCTGATCACTCAAAATAATATCAACAAGAACCAGTCTGCGCTGTCGAGTTCTATC
 GAGCGTCTGTCTTCTGGCTTGCGTATTAAACAGCGCGAAGGATGACGCCGAGGTCTAGGCG
 ATTGCTAACCGTTTTACTTCTAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAAC
 GACGGTATTTCTGTTGCACAGACCACTGAAGGCGCGTGTCCGAAATCAACAACAACTTA
 CAGCGTGTGCGTGAAGTACCCTTCAGGCAACCACCGGTACCAACTCCCAGTCTGACCTG
 GACTCTATCCAGGACGAAATTAAATCCCGTCTGGACGAAATTGATCGCGTATCCGGTCAG
 ACCCAGTTCAACGGCGTGAACGTGCTGGCAAAAGACGGTTCATGAAAATTTCAGGTTGGC
 GCGAACGATGGCCAGACCATCACTATCGACCTGAAGAAGATTGACTCTTCTACCTTGAAC
 CTGACAGGTTTTAACGTTAACGGTTCTGGTTCTGTGGCGAATACTGCAGCAACTAAAGCT
 GATTTAACCGCTGCTCAACTCTCTGCACCGGGTGCAGCAGACGCAAATGGTACAGTTACT
 TATACTGTGCTGCTGGTTATAAAGAATCCACTGCTGCAGATGTTATTGCTAGCATCAAA
 GACGGCAGTGTCCGACTTCTGCAATTACTGCAACCATTAATAATGGCTTCGGTGATTCC
 AGTGCGCTGACTTCCAATGACTATACCTTATGACCCAGCAAAAGGCGACTTCACTTACGAC
 GTAGCTTCAAGCGCCAATAATACTGCTGCCCAGGTTGAGTCTTCTGACGCCGAAAGCA
 GGTGATACCGCAAATCTGAAAGTAACCGTTGGTACGACATCGGTTGATGTCGTTCTGGCC
 AGTGATGGTAAGATTACAGCAAAAGATGGTTCTGCATTATATATCGACAGTACAGGTAAC
 CTGACTCAGAACAGTGTGGCTTGACCTCTGCTAAACTGGCTACTCTGACTGGCCTTCAG
 GGCTCTGGTGTGCTTCAACCATCACTACTGAAGATGGCACTAATATTGATATTGCTGCT
 AACGGTAATATTGGTCTGACCGGTGTTCTGATCAGTGCTGATTCTCTGCAGTCAGCGACT
 AAATCTACGGGCTTTACTGTTGGTACTGGCGCTACAGGTCTGACCGTAGGTACTGATGGT
 AAAGTGACTATCGGCGGGACTACTGCTCAGTCCTACACCAGCAAAGATGGTTCCCTGACT
 ACTGATAACACCACTAAACTGTATCTGCAGAAAGATGGCTCTGTAACCAACGGTTCAGGT
 AAAGCGGTCTATGTAGAAGCGGATGGTGATTTCACTACCGACGCTGCAACCAAAGCCGCA
 ACCACCACCGATCCGCTGAAAGCCCTGGATGAGGCAATCAGCCAGATCGATAAGTTCCGT
 TCATCCCTGGGTGCTATCCAGAACCGTCTGGATTCCGCGGTACCAACCTGAACAACACC
 ACTACCAACCTGTCTGAAGCGCAGTCCCGTATTAGGACGCGGACTATGCGACCGAAGTG
 TCCAACATGTCGAAAGCGCAGATCATTAGCAGGCCGGTAACTCCGTGCTGGCAAAAGCC
 AACCAGGTACCGCAACAGGTTCTGTCTCTGCTGCAGGGCTAA

Figure 20

ATGGCACAAGTCATTAATACCAACAGCCTCTCGCTGATCAC
 TCAAAATAATATCAACAAGAACAGTCTGCGCTGTCGAGTTCATCGAGCGTCTGTCTTC
 TGGCTTGCGTATTAAACAGCGCGAAGGATGACGCCGAGGTGAGGCGATTGCTAACCGTTT
 TACTTCTAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAACGACGGTATTTCTGT
 TGCACAGACCACTGAAGGCGCGCTGTCCGAAATCAACAACAACCTTACAGCGTGTGCGTGA
 ACTGACCGTTTCAAGCAACCACCGGTACCAACTCCCAGTCTGACCTGGAATCTATCCAGGA
 CGAAATTAAATCCCGTCTGGACGAAATTGATCGCGTATCCGGTCAGACCCAGTTCAACGG
 CGTGAACGTGCTGGCAAAAGACGGTTCCATGAAAATTGAGTTGGCGCAACGATGGCCA
 GACCATCACTATCGACCTGAAGAAGATTGACTCTTCTACCTTGAACCTGACAGGTTTAA
 CGTTAACGGTTCTGTTCTGTGGCGAATACTGCAGCAACTAAAGCTGATTTAACCGCTGC
 TCAACTCTCTGCACCGGGTGCAGCAGACGCAATGGTACAGTTACTTATACTGTCAGTGC
 TGGTTATAAAGAATCCACTGCTGCAGATGTTATTGCTAGCATCAAAGACGGCAGTGCTCC
 GACTTCTGCAATTACTGCAACCATTAATAATGGCTTCGGTGATTCCAGTGCGCTGACTTC
 CAATGACTATACCTATGACCCAGCAAAAGGCGACTTCACTTACGACGTAGCTTCAAGCGC
 CAATAATACTGCTGCCCAGGTTGAGTCTTCTCTGACGCCGAAAGCAGGTGATACCGCAA
 TCTGAAAGTAACCGTTGGTACGACATCGGTTGATGTCGTTCTGGCCAGTGATGGTAAGAT
 TACAGCAAAGATGGTTCTGCATTATATATCGACAGTACAGGTAACTGACTCAGAACAG
 TGCTGGCTTGACCTCTGCTAACTGGCTACTCTGACTGGCCTTCAGGGCTCTGGTGTTGC
 TTCAACCATCACTACTGAAGATGGCACTAATATTGATATTGCTGCTAACGGTAATATTGG
 TCTGACCGGTGTTGCTATCAGTGCTGATTCTCTGCAGTCAGCGACTAAATCTACGGGCTT
 TACTGTTGGTACTGGCGCTACAGGTCTGACCGTAGGTACTGATGGTAAAGTGAATATCGG
 CGGGACTACTGCTCAGTCCCTACACCAGCAAAGATGGTTCCCTGACTACTGATAACACCAC
 TAACTGTATCTGCAGAAAGATGGCTCTGTAACCAACGGTTGAGTTAAAGCGGTCTATGT
 AGAAGCGGATGGTGATTTCACTACCGACGCTGCAACCAAAGCCGCAACCACCACCGATCC
 GCTGAAAGCCCTGGATGAGGCAATCAGCCAGATCGATAAGTTCCGTTTATCCCTGGGTGC
 TATCCAGAACCGTCTGGATTCCGCGGTACCAACCTGAACAACCACTACCAACCTGTC
 TGAAGCGCAGTCCCGTATTGAGGACGCCGACTATGCGACCGAAGTGTCCAACATGTCGAA
 AGCGCAGATCATTCAGCAGGCCGGTAACCTCCGTGCTGGCAAAAGCCAACCAGGTACCGCA
 ACAGGTTCTGTCTCTGCTGCAGGGCTAA

Figure 21

GCGCTGTCGACTTCTATCGAGCGCCTCTCTTCTGGTCTGCGTATTAACAGCGCTAAA
GATGACGCTGCGGGCCAGGCGATTGCTAACCCTTCACTTCTAACATCAAAGGTCTGACT
CAGGCCGCACGTAACGCCAACGACGGTATTTCTCTGGCGCAGACGGCTGAAGGCGCGCTG
TCAGAGATTAACAACAACCTTGCAGCGTATTCGTGAACGACCGTTCAGGCCTCTACCGGC
ACGAACTCTGATTCCGACCTGTCTTCTATTTCAGGACGAAATCAAATCCCGTCTTGATGAA
ATTGACCGTGTATCTGGTCAGACCCAGTTCAACGGTGTGAACGTGCTGTCGAAAAACGAT
TCGATGAAGATTGAGTTGGTGCCAATGATAACCAGACGATCAGCATTGGCTTGCAACAA
ATCGACAGTACCACTTTGAATCTGAAAGGATTTACCGTGTCCGGCATGGCGGATTTTCAGC
GCGGCGAAACTGACGGCTGCTGATGGTACAGCAATTGCTGCTGCGGATGTCAAGGATGCT
GGGGGTAAACAAGTCAATTTACTGTCTTACACTGACACCGCGTCTAACAGTACTAAATAT
GCGGTCTGTTGATTCTGCAACCGGTAAATACATGGCAGCCACTGTAGTCATTACCAGTACG
GCGGCGGCGGTAACTGTTGGTGCAACGGAAGTGGCGGGAGCCGCTACAGCCGAACCGTTA
AAAGCACTGGATGCCGCAATCGCTAAAGTCGACAAATTCCGCTCCTCCCTCGGTGCCGTT
CAAAACCGTCTGGATTCTGCGGTCACCAACCTGAACAACACCACCACCAACCTGTCTGAA
GCGCAGTCCCGTATTTCAGGACGCCGACTATGCGACCGAAGTGTCCAACATGTCGAAAGCG
CAGATTATCCAGCAGGCG

Figure 22

ATGGCACAAAGTCATTAATACCAACAGCCTCTCGCTGATCACTCAAAATA
 ATATCAACAAGAACCAGTCTGCGCTGTTCGAGTTCTATCGAGCGTCTGTCTTCTGGCTTGC
 GTATTAACAGCGCGAAGGATGACGCCGAGGTCAGGCGATTGCTAACCGTTTTACTTCTA
 ATATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAATGACGGTATTTCTGTTGCACAGA
 CCACTGAAGGCGCGCTGTCCGAAATCAACAACAACCTTACAGCGTATTCGTGAACTGACGG
 TTCAGGCCACTACAGGGACTAACTCCGATTCTGACCTGGACTCCATCCAGGACGAAATCA
 AATCTCGTCTGGACGAAATTGACCGCGTATCCGGTCAGACCCAGTTCAACGGCGTGAACG
 TGCTGTCCAAAGATGGTTCAATGAAAATTACAGGTCGGCGCAAATGATGGTGAAACCATCA
 CGATTGATCTGAAGAAAATTGACTCTGATACGCTGAATCTGGCTGGTTTTAACGTGAATG
 GCGAAGGTGAAACAGCCAATACTGCTGCAACACTTAAAGATATGGTTGGTTTTAAACTCG
 ATAATACGGGGGTCACTACAGCTGGAGTTAATAGATATATTGCTGACAAAGCCGTCGCAA
 GTAGCACGGATATTTTGAATGCGGTAGCTGGTGGTGGATGGCAGTAAAGTTTCCACGGAGG
 CAGATGTTGGTTTTGGTGCAGCTGCCCCCTGGTACGCCAGTGGAATATACTTATCATAAAG
 ATACTAACACATATACGGCTTCTGCTTCAGTTGATGCGACTCAACTGGCGGCATTCTCTGA
 ATCCTGAAGCGGGTGGTACCCTGCTGCAACAGTAAGTATTGGCAACGGTACAACAGCTC
 AAGAGCAAAAAGTCATTATTGCTAAAGATGGTTCTTTAACTGCTGCTGATGACGGTGCCG
 CTCTCTATCTTGATGATACTGGTAACTTAAGTAAAACTAACGCAGGCACTGATACTCAAG
 CTAAACTGTCTGACTTAATGGCAAACAATGCTAATGCCAAAACAGTCATTACAACAGATA
 AAGGTACATTTACTGCTAATACGACAAAGTTTGATGGGGTAGATATTTCTGTTGATGCTT
 CAACGTTTGCTAACGCCGTTAAAAATGAGACTTACACTGCAACTGTTGGTGTAACTTTAC
 CTGCGACATATACAGTCAATAATGGCACTGCTGCATCAGCGTATTTAGTCGATGGAAAAG
 TGAGCAAACTCCTGCCGAGTATTTTGCTCAAGCTGATGGCACTATTACTAGTGGTGAAA
 ATGCGGCTACCAGTAAAGCTATCTATGTAAGTGCCAATGGTAACTTAACGACTAATACAA
 CTAGTGAATCTGAAGCTACTACCAACCCGCTGGCAGCATTGGATGACGCTATCGCGTCTA
 TCGACAAATTCCGTTCTTCCCTGGGTGCTATCCAGAACCGTCTGGATTCCGCAGTCACCA
 ACCTGAACAACACCACTACCAACCTGTCTGAAGCGCAGTCCCGTATTCAGGACGCCGACT
 ATGCGACCGAAGTGTCACACATGTGAAAGCGCAGATCATTCAGCAGGCCGGTAACTCCG
 TGCTGGCAAAAGCCAACCAGGTACCGCAGCAGGTTCTGTCTCTGCTGCAGGGTTAA

Figure 23

ATGGCACAAAGTCATTAATACCAACAGCCTCTCGCTGATCACTCAAAATAATAT
 CAACAAGAACCAGTCTGCGCTGTCGAGTTCTATCGAGCGTCTGTCTTCTGGCTTGCGTAT
 TAACAGCGCGAAGGATGACGCCGAGGTGAGGCGATTGCTAACCGTTTACTTCTAACAT
 TAAAGGCCTGACTCAGGCTGCACGTAACGCCAACGACGGTATTTCTGTTGCGCAGACCAC
 TGAAGGCGCGCTGTCCGAAATTAACAACAACCTTACAGCGTATTCGTGAACTGACGGTTCA
 GGCGACGACCCGAACTAACTCCACCTCTGACCTGGACTCCATCCAGGACGAAATCAAATC
 CCGTCTTGACGAAATTGACCGCGTATCTGGTCAGACCCAGTTCAACGGCGTGAACGTGCT
 GTCTAAAGATGGCTCGATGAAAATTGAGGTCGGCGCGAACGATGGCGAAACGATTACTAT
 TGATCTGAAGAAAATTGACTCTGATACGCTGAATCTGGCTGGTTTTAACGTTAACGGTAA
 AGGTTCTGTAGCGAATACCGCTGCGACTACAGATAATCTGACATTGGCTGGTTTTACAGC
 GGGTACTAAAGCTGCTGATGGCACCGTAACCTATAGCAAAAATGTCCAGTTTGCCGCCGC
 GACTGCAAGCAATGTACTGGCTGCTGCTAAAGATGGCGACGAAATTACGTTGCTGGTAA
 TAACGGCACAGGTATAGCTGCAACTGGGGGACTTATACTTATCATAAGGACTCTAACTC
 ATACAGCTTTAGCGCAACGGCTGCATCTAAAGATTCTCTGTTGAGCACACTGGCACCAAA
 CGCTGGCGATACATTTACCGCTAAAGTGACTATTGGTTCTAAATCGCAAGAAGTTAACGT
 TAGCAAAGATGGTACGATTACATCCAGCGATGGTAAGGCGCTGTATTTAGATGAGAAGGG
 CAACCTGACCCAAACAGGTAGTGGCACAACCAAGCTGCAACCTGGGATAACCTGATGGC
 CAATACAGATACTACAGGCAAAGATGCCTATGGTAACTCTGCGGCAGCAGCTGTTGGGAC
 AGTAATCGAAGCAAAAGGAATGACCATCACTTCTGCTGGTGGTAATGCTCAGGTGTTAAA
 AGACGCGGCTTATAATGCCGCATATGCGACCTCAATTACTACTGGTACTCCGGGTGATGC
 GGGAGCCGCGGAGCCGCTGCAACTGCGGGTAATGCCGCGGTGGGAGCGCTGGGCGCAAC
 GGCAGTTGATAATACCACGGCAGATGTTGCCGATATCTCTATCTCAGCTTCGCAAATGGC
 GAGCATCCTTCAGGATAAAGATTTACCTTAAGTGATGGTAGTGATACTTACAACGTGAC
 CAGCAATGCTGTCACTATCAATGGCAAAGCAGCAAACATTGATGACAGCGGCGCAATCAC
 AGACCAAACCAGTAAAGTTGTCAATTATTTGCTCATACTAACGGTAGCGTGACTAACGA
 TACAGGCTCCACTATTTATGCGACAGAAGATGGTAGCCTGACCACCGATGCAGCAACCAA
 AGCCGAAACCACCGCCGATCCCCTGAAAGCTCTGGACGAAGCCATCAGCTCCATCGACAA
 ATTCCGCTCCTCCCTCGGTGCGGTGCAAAACCGTCTGGATTCCGCGGTACCAACCTGAA
 CAACACCACCACCAACCTGTCTGAAGCGCAGTCCCGTATTGAGGACGCCGACTATGCGAC
 CGAAGTGTCACATGTGAAAGCGCAGATTATCCAGCAGGCCGGTAACTCCGTGCTGGC
 AAAAGCTAACAGGTACCACAGCAGGTTCTGTCTCTGCTGCAGGGTTAA

Figure 24

ATGGCACAAGTCATTAATACCAACAGCCTCTCGCTG
 ATCACTCAAAATAATATCAACAAGAACCAGTCTGCGCTGTCGAGTTCTATCGAGCGTCTG
 TCTTCTGGCTTGCGTATTAAACAGCGCGAAGGATGACGCCGAGGTGAGGCGATTGCTAAC
 CGTTTTACTTCTAACATTAAAGGCCCTGACTCAGGCGGCCCGTAACGCCAACGACGGTATT
 TCTGTTGCGCAGACCACCGAAGGCGCGCTGTCCGAAATTAACAACAACCTTACAGCGTGTG
 CGTGAGCTGACTGTTTACGGCGACCACCGGTACCAACTCCAGTCTGATCTGGACTCTATC
 CAGGACGAAATCAAATCCCGTCTGGACGAAATTGACCGCGTATCCGGTCAGACCCAGTTC
 AACGGCGTGACGTGCTGGCAAAAGACGGTTCCATGAAAATTGAGGTTGGCGCGAATGAT
 GGCCAGACCATCACTATCGACCTGAAGAAGATTGACTCTTCTACGTTGAACTGACTGGT
 TTTAACGTGAATGGTTCTGGTTCTGTGGCGAATACTGCGGCGACTAAAGCGGATTTGGCT
 GCTGCTGCAATTGGTACCCCTGGGGCAGCAGATTCTACAGGTGCCATTGCTTACACAGTA
 AGTGCTGGGCTGACTAAACTACAGCCGCAGATGTACTGTCTAGCCTCGCTGATGGTACG
 ACTATTACAGCCACAGGCGTGAAAAATGGCTTTGCTGCAGGAGCCACTTCCAATGCCTAT
 AAACCTTAACAAAGATAATAATACATTTACTTATGACACGACTGCTACGACAGCTGAGCTG
 CAGTCTTACCTGACTCCGAAAGCGGGCGACACTGCAACATTCAGTGTGAAATTGGTGGT
 ACTACACAAGACGTCGTGCTGTCCAGTGATGGCAAACTCACTGCTAAGGATGGCTCTAAG
 CTTTACATTGATACAACCTGGTAATTTAACTCAGAATGGTGGTAATAACGGTGTGGAACA
 CTCGCGGAAGCGACTCTGAGTGGTTTAGCTCTGAACAAAAATGGTTTAAACGGCTGTTAAA
 TCCACAATTACTACAGCTGATAACACTTCGATTGTACTGAATGGTTCAAGCGATGGTACT
 GGTAAATGCTGGTACTGAAGGTACGATTGCTGTTACAGGCGCTGTAATTAGTTCAGCTGCT
 CTGCAATCTGCAAGCAAAACGACTGGTTTCACTGTTGGTACAGTAGACACAGCTGGTTAT
 ATCTCTGTAGGTACTGATGGGAGTGTTTCAAGGCATATGATGCTGCGACTTCTGGCAACAAA
 GCTTCTTACACCAACACTGACGGTACACTGACTACTGATAACACCACTAAACTGTATCTG
 CAGAAAGATGGCTCTGTAAACCAACGGTTTCAAGTAAAGCGGTCTATGTAGAAGCGGATGGT
 GATTTCACTACCGACGCTGCAACCAAGCCGCAACCACCACCGATCCGCTGGCCGCTCTG
 GATGACGCAATCAGCCAGATCGACAAGTTCCGTTTCATCCTTGGGTGCTATCCAGAACCGT
 CTGGATTCTGCAGTCACCAACCTGAACAACACCACCACCAACCTGTCTGAAGCGCAGTCC
 CGTATTCAAGGACGCCGACTATGCGACCGAAGTGTCCAATATGTGCAAAAGCGCAGATCATC
 CAGCAGGCCGGTAACTCCGTGCTGGCAAAAGCCAACCAGGTACCGCAGCAGGTTCTGTCT
 CTGCTGCAGGGTTAA

Figure 25

AACAAATCTCAGTCTTCTCTGAGCTCCGCCATTGAA
CGTCTCTCTTCTGGCCTGCGTATTAACAGTGCTAAAGATGACGCAGCAGGTCAGGCGATT
GCTAACCGTTTTACAGCAAATATTAAAGGTCTGACTCAGGCTTCCCGTAACGCGAATGAT
GGTATTTCTGTTGCGCAGACCACTGAAGGTGCGCTGAATGAAATTAACAACAACCTGCAG
CGTATTCGTGAACTTTCTGTTCAAGCACTAACGGTACTAACTCTGACAGCGATCTTTCT
TCTATCCAGGCTGAAATTACTCAACGTCTGGAAGAAATTGACCGTGTATCTGAGCAAACCT
CAGTTTAACGGCGTGAAAGTCCTTGCTGAAAATAATGAAATGAAAATTCAGGTTGGTGCT
AATGATGGTGAAACCATCACTATCAATCTGGCAAAAATTGATGCGAAAACCTCTCGGCCTG
GACGGTTTTAATATCGATGGCGCGCAGAAAGCAACCGGCAGTGACCTGATTTCTAAATTT
AAAGCGCAGGTACTGATAATTATCAAATTAACGGTACTGATAACTATACTGTTAATGTA
GATAGTGGCGTAGTACAGGATAAAGATGGCAAACAAGTTTATGTGAGTACTGCGGATGGT
TCACTTACGACCAGCAGTGATACTCAATTCAAGATTGATGCAACTAAGCTTGCACTGGCT
GCTAAAGATTTAGCTCAAGGGAATAAGATTGTCTACGAAGGTATCGAATTTACAAATACC
GGCACTGTGCGCTATAGATGCCAAAGGTAATGGTAAATTAACCGCCAATGTTGATGGTAAG
GCTGTTGAATTCATATTTTCGGGGAGTACTGATACATCAGGTACTAGTGCAACCGTTGCC
CCTACGACAGCCCTATACAAAAATAGTGCAGGGCAATTGACTGCAACAAAAGTTGAAAAT
AAAGCAGCGACACTATCTGATCTTGATCTGAACGCTGCCAAGAAAACAGGAAGCACGTTA
GTTGTTAACGGTGCAACTTACGATGTTAGTGCAGATGGTAAAACGATAACGGGAGACTGCT
TCTGGTAACAATAAAGTCATGTATCTGAGCAAATCAGAAGGTGGTAGCCCGATTCTGGTA
AACGAAGATGCAGCAAAATCGTTGCAATCTACCACCAACCCGCTCGAACTATCGACAAA
GCATTGGCTAAAGTTGACAATCTGCGTTCTGACCTCGGTGCAGTACAAAACCGTTTCGAC
TCTGCCATACCAACCTTGGCAACACCGTAAACAACCTGTCTTCTGCCCCTAGCCGTATC
GAAGATGCTGACTACGCGACCGAAGTGTCTAACATGTCTCGTGCGCAGATCCTGCAACAA
GCGGGTACCTCTGTTCTGGCACAG

Figure 26

ATGGCACAAGTCATTAATACCAACAGCCTCTCGCTGATCACTCAAAATA
ATATCAACAAGAACCAGTCTGCGCTGTGAGTTCTATCGAGCGTCTGTCTTCTGGCTTGC
GTATTAACAGCGCGAAGGATGACGCAGCGGGTCAGGCGATTGCTAACCGTTTTACTTCTA
ACATTAAAGGCCTGACTCAGGCGGCACGTAACGCCAACGACGGTATCTCTCTGGCGCAGA
CCACCGAAGGTGCGCTGTCTGAAATCAACAACAACTTACAGCGTGTACGTGAACTGACCG
TTCAGGCAACCACCGGTACTAACTCCGACTCCGACCTGGCTTCTATTACAGACGAAATCA
AATCCCGTCTGGATGAAATTGACCGCGTATCTGGTCAGACTCAGTTCAACGGCGTGAACG
TGCTGGCAAAGACGGTTCATGAAAATTACGGTAGGTGCTAACGACGGCCAGACTATCA
CTATTGACCTGAAAAAATCGACTCTGATACTCTGGGCCCTGAATGGTTTAAACGTGAATG
GTTCTGGGACGATTACCAACAAAGCAGCAACTGTCAGTGATGTTACTCGCGCAGGCGGTA
CATTGGTGAATGGTGCCATGATATAAAAACCACTAACACAGCGCTGACTACAACCTGATG
CCTTCGCGAAATTGAATGATGGTGATGTTGTTACTATCAATAATGGTAAGGATACTGCCT
ATAAATATAATGCTGCTACAGGTGGGTTTACGACGGATGTCTCCATCTCCGGGGATCCTA
CCGCTGCTGACGCTACTGCTAATAAACTGCCCCGTGATGCACTTGCGGCGTCTTTACATG
CTGAGCCGGGTAAACTGTTAATGGTTCCTGGACTACGAATGATGGTACGGTAAATTTG
ATACCGATGCCGATGGTAAGATTTCTATTGGTGGTGTGCTGCTTATGTAGATGCAGCAG
GCAACCTGACCACTAACGCAGCAGGTATGACGACTCAAGCAACAACCTACCGATTTGGTTA
CTGCTGCTGCATCTGCTACTGGTAAGGGTGGATCCCTGACCTTTGGTGACACGACGTATA
AAATTGGTCAGGGTACGGCTGGGGTTGATCCTGATGACGCTTCAGATGATGTACTGGGCA
CCATTTCTTACTCTAAATCAGTAAGCAAGGATGTTGTTCTTGCTGATACTAAAGCAACTG
GTAACACGACAACAGTTGATTTCAACTCCGGTATCATGACTTCAAAGGTTAGTTTCGATG
CAGGTACATCAACTGATACATTCAAAGATGCAGATGGTGCTATCACCAAAACCTAAAGAAT
ACACCACTTCTTATGCTGTAAATAAAGATACTGGTGAAGTTACCGTTGCTGATTATGCTG
CGGTAGATAGCGCCGATAAGGCTGTTGATGATACTAAATATAAACCGACTATCGGCGCGA
CAGTTAACCTGAATTCTGCAGGTAAATTGACCACTGATACCACCACTGCAGGCACAGCAA
CCAAAGATCCTCTGGCTGCCCTGGACGCTGCTATCAGCTCCATCGACAAATTCCGTTTAT
CCCTGGGTGCTATCCAGAACCGTCTGGATTCCGCAGTCACCAACCTGAACAAACACCACTA
CCAACCTGTCCGAAGCGCAGTCCCGTATTACAGGACGCGACTATGCGACCGAAGTGTCCA
ACATGTCGAAAGCGCAGATTATCCAGCAGGCCGGTAACTCCGTGCTGGCAAAGCCAACC
AGGTACCGCAGCAGGTTCTGTCTCTGCTACAGGGTTAA

Figure 27

AACAAAAACCAGTCTGCGCTGTGCGACTTCTATC
 GAGCGCCTTTCTTCTGGTCTGCGTATTAACAGCGCTAAAGATGACGCTGCGGGCCAGGCG
 ATTGCTAACCGCTTCACCTTCTAACATCAAAGGTCTGACTCAGGCCGCACGTAACGCCAAC
 GACGGTATTTCTCTGGCGCAGACCACTGAAGGCGCGCTGTCTGAGATTAACAACAACCTTG
 CAGCGTGTGCGTGAGTTGACTGTACAGGCGACGACCGGGACTAACTCTGATTCTGACCTG
 TCTTCTATCCAGGATGAAATCAAATCCCGTTTAAGCGAAATTGACCGTGATCTGGTCAG
 ACTCAGTTTAACGGCGTGAACTGACTGGCTAAGAATGACACCTGTCTATTCAGGTAGGT
 GCAAATGACGGTCAGACTATCAATATTGACCTGCAGCAAATCGATTCTCATACACTGGGT
 CTGGATGGTTTCAGCGTTAAAAATAATGATGCAGTGAAAACCACTGCTGCCGTGAATACT
 CTTGGGGGGGGGCGAGTTCTGTTGCTGTCGACTTCGCAACAACCACTTTGACTGCTATC
 ACTGGTCTCGGTAGCGGTGCTATCAGCGAAATTGCTAAAGACGATAATGGTGATTACTAC
 GCGCATGTCACAGGGACTACGGGTAATACTGCTGATGGTTACTATGCTGTGATATCGAC
 AAGGCTACCGGTGAGGTGCTCTGAAAGATGGTAACGTAGATACACCGACAGGTACGCCA
 ACGACGACAAGCACATATGACTTCACAGACGCTGGTCAAACCGTTTCCTTTGGCACTGAT
 GCTGCAACAGCCGGTATCAGCACTGGTGCTTCTCTCGTTAACTTCAGGATGAGAAAGGC
 AATGATACTGCTACTTATGCAATCAAAGCACAAAGATGGCAGCCTGTATGCCGCCAACGTT
 GATGAGGCTACCGGTAAAGTCACTGTCAAACCGCCAGCTATACTGATGCTGACGGCAAA
 GCAGTGACCGATGCCGCTGTAAACTGGGTGGTGACAATGGCACAACCGAAATTGTTGTC
 GATGCTGCGTCAGGTAAACTTACGATGCTGGTGCACTGCAAAACGTTGATCTCTCCAGT
 GCAACCAACACGGTAACCGCAATCCCGAACGGTAAACACGCTCTCCGCTGGCTGCCCTT
 GACGACGCAATCAGCCAGATCGACAAATTCGCTCCTCCCTCGGTGCGGTGCAGAACCGT
 CTGGATTCCGCGGTACCAACCTGAACAACCACTACCAACCTGTCTGAAGCGCAGTCC
 CGTATTGAGGACGCTGACTATGCGACCGAAGTATCCAACATGTGAAAGCGCAGATCATC
 CAGCAGGCAGGTAACCTCCGTGCTGTCCAAA

Figure 28

GCGCTGTCGACTTCTATCGAGCGCCTCTCTTCTGGTCTGCGCATTAACAGCGCTAAAG
ATGACGCTGCGGGCCAAGCGATTGCTAACCGCTTCACTTCTAACATCAAAGGTCTGACTC
AGGCCGCACGTAACGCCAACGACGGTATTTCTCTGGCGCAGACCACTGAAGGCGCACTGT
CTGAAATCAACAACAACCTTGACGCGTGTTCGTGAACTGACCGTTCAGGCCACTACCGGTA
CTAACTCTGATTCTGACCTGTCTTCAATACAGGACGAAATCAAATCCCGTCTCGATGAAA
TTGACCGCGTATCCGGTCAGACTCAGTTCAACGGCGTTAATGTTCTTTCCAAAGATGGTT
CAATGAAAATTCAGGTTGGTGCGAATGATGGTCAAATCTATCTCCATCGATCTGAAGAAAA
TTGATTCTTCAACTTTGGGGCTGAATGGCTTCTCAGTTTCTAAAAACTCTCTTAATGTCA
GCAATGCTATCACATCTATCCCGCAAGCCGCTAGCAATGAACCTGTTGATGTTAACTTCG
GTGATACTGATGAGTCTGCAGCAATCGCAGCCAAATTGGGGGTTTCCGATACGTCAAGCC
TGTCGCTGCACAACATCCTTGATAAAGATGGTAAGGCAACAGCTGATTATGTTGTTCACT
CAGGTAAAGACTTCTATGCTGCTTCTGTTAATGCCGCTTCAGGTAAAGTAACCTTAAACA
CCATTGATGTTACTTATGATGATTATGCGAACGGTGTGACGATGCCAAGCAAACAGGTC
AGCTGATCAAAGTTTCAGCAGATAAAGACGGCGCAGCTCAAGGTTTGTACACTTCAAG
GCAAAAACTATTCTGCTGGTGATGCGGCAGACATTCTTAAGAATGGAGCAACAGCTCTTA
AGTTAACTGATCTGAATTTAAGTGATGTTACTGATACTAATGGTAAGGTAACCACAACCTG
CGACTGAGCAATTTGAAGGTGCTTCAACTGAGGATCCGCTGGCGCTTCTGGATAAAGCTA
TTGCATCAGTCGACAAATTCGGTCTTCTCTAGGTGCCGTGCAGAACCGTCTCGATTCCG
CTATACCAACCTGAACAACACCACCACCAACCTGTCTGAAGCGCAGTCCCGTATTCAAG
ACGCCGACTATGCGACCGAAGTGTCCAACATGTCGAAAGCGCAGATCATCCAGCAGGCA

Figure 29

ATGGCACAAGTCATTAATACCAACAGCCTCTCG
 CTGATCACTCAAAATAATATCAACAAGAACCAGTCTGCGCTGTCGAGTTCTATCGAGCGT
 CTGTCTTCTGGCTTGCCTATTAAACAGCGCGAAGGATGACGCCGAGGTGAGGCGATTGCT
 AACCGTTTTACTTCTAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAACGACGGT
 ATTTCTGTTGCACAGACCACTGAAGGCGCGCTGTCCGAAATCAACAACAATTACAGCGT
 ATTCGTGAAGTACGGTTTACGGCCACTACAGGGACTAACTCCGATTCTGACCTGGACTCC
 ATCCAGGACGAAATCAAATCTCGTCTGGACGAAATTGACCGCGTATCTGGTCAGACCCAG
 TTCAACGGCGTGAACGTGCTGTCTAAAGATGGCTCGATGAAAATTGAGTTCGGCGCGAAC
 GATGGCGAAACGATTACTATTGATCTGAAGAAAATTGACTCTGATACGCTAAATCTGGCT
 GGTTTTAACGTGAATGGTGTGGCTCTGTTGATAATGCCAAGGCGACTGGCAAAGATCTT
 ACTGATGCTGGTTTTACGGCAAGCGCAGCTGATGCTAATGGCAAATCACTTATACCAA
 GACACCGTTACTAAATTGACAAAGCGACAGCGGCTGATGTATTGGGCAAAGCGGCTGCT
 GCGGATAGCATTACCTATGCGGGCACTGATACTGGCTTAGGAGTCGCTGCTGATGCCTCG
 ACTTACACCTACATGCAGCCAATAAGTCTTACACTTTTGATGCTACTGGTGTGCGCAAG
 GCGGATGCTGGAACGGCACTGAAAGGGTACTTAGGCGCATCTAACACCGGTAATAAT
 ATCGGTGGTACCGAGCAAGAAGTTAACAATTGCCAAAGATGGCTCCATCACCAGATACCAAT
 GCGGATGCGCTGTATCTCGATAGTACCGGCACTTAACCAAAAATACCGCAATTTGGGG
 GCTGCTGATAAAGCAACTGTAGATAAACTGTTTGCTGGTGTCTCAGGATGCAACGATCACC
 TTCGATAGCGGCATGACAGCTAAATTCGATCAAATGCTGGTACCGTTGATTTCAAAGGC
 GCGTCTATTTCTGCTGATGCAATGGCATCAACCTTAAATAATGGTTCTTATACAGCCAAC
 GTAGGTGGTAAGGCTTATGCCGTAACCGCTGGCGCAGTTCAGACAGGTGGCGCAGATGTG
 TATAAAGATACCACTGGCGCACTGACGACTGAAGATGACGAAACCGTTACCGCGACCTAC
 TACGGTTTTGCTGATGGTAAAGTTTCTGACGGTGAAGGTTCTACTGTCTATAAAGCTGCT
 GATGGTTCCATCACTAAAGATGCGACTACCAAGTCTGAAGCAACCACTGACCCTCTGAAA
 GCCCTTGACGACGCAATCAGCCAGATCGACAAATCCGCTCCTCCCTCGGTGCCGTTCAA
 AACCGTCTGGATTCCGCCGTACCAACCTGAACAACACCACTACCAACCTGTCTGAAGCG
 CAGTCCCGTATTACGAGACGCCGACTATGCGACCGAAGTGTCCAACATGTGAAAGCGCAG
 ATCATTCAGCAGGCCGGTAACCTCCGTGCTGGCAAAGCCAACAGGTACCGCAGCAGGTT
 CTGTCTCTGCTGCAGGGTTAA

Figure 30

AACAAATCTCAGTCTTCTCTTAGCTCTGCTATTGA
 GCGTCTCTCTTCTGGCCTGCGTATTAAACAGTGCTAAAGATGACGCAGCAGGTCAGGCGAT
 TGCTAACCGTTTTACGGCAAATATTAAAGGTCTGACTCAGGCTTCCCGTAACGCGAATGA
 TGGTATTTCTGTTGCGCAGACTACTGAAGGTGCGCTGAATGAAATTAACAACAACCTGCA
 GCGTGACGTGAACTGACTGTTTCAAGCAACTAACGGTACTAACTCTGACAGCGATCTTTC
 TTCTATTTCAGGCAGAAATTACTCAACGTCTGGAAGAAATTGACCGTGTATCTGAGCAAAC
 TCAGTTTAAACGGCGTGAAAGTCCTTGCCGAAAATAATGAAATGAAAATTGAGGTTGGTGC
 TAATGATGGGGAAACCATCACTATCAATCTGGCAAAAATTGATGCGAAAACCTCTCGGCCT
 GGACGGCTTTAATATCGATGGCGCGCAGAAAGCAACTGGCAGTGACCTGATTTCTAAAT
 TAAAGCGACAGGTACTGATAATTATCAAATTAACGGTACTGATAACTATACTGTTAATGT
 AGATAGTGGAGCAGTTCAAATGAGGATGGTGACGCAATTTTGTAGCGCTACCGATGG
 TTCTCTGACTACTAAGAGTGATACAAAAGTCGGTGGTACAGGTATTGATGCGACTGGGCT
 TGCAAAAGCCGCGAGTTTCTTTAGCTAAAGATGCCTCAATTAAATACCAAGGTATTACTTT
 CACCAACAAAGGCACTGATGCATTTGATGGCAGTGGTAACGGCACTCTAACCGCTAATAT
 TGATGGCAAAGATGTAACCTTTACTATTGATGCGACAGGGAAGGACGCAACATTAAAAAC
 GTCTGATCCTGTTTACAAAAATAGTGCAGGTGAGTTCACTACAATAAGGTTGAAAACAA
 AGCCGCTACAGCATCGGATCTGGACTTAAATAACGCTAAAAAAGTGGGTAGTTCTTTAGT
 TGTAAATGGCGCTGATTATGAAGTTAGCGCTGATGGTAAGACAGTAAGTGGGCTTGGCAA
 AACTATGTATCTGAGCAAATCAGAAGGTGGTAGCCGATTCTGGTAAAAGAAGATGCAGC
 AAAATCGTTGCAATCTACTACCAACCCGCTCGAAACCATCGACAAGGCATTGGCTAAAGT
 TGACAATCTGCGTTCTGACCTCGGTGCGAGTACAAAACCGTTTCGACTCTGCTATCACCAA
 CCTTGGCAACACCGTAAACAACCTGTCTTCTGCCCGTAGCCGTATCGAAGATGCTGACTA
 CGCGACCGAAGTGTCTAACATGTCTCGTGCGCAGATCCTGCAACAAGCGGGTACCTCTGT TCTGGCGCAG

Figure 31

ATGGCACAAGTCATTAATACCAACAGCCTCTCGCTGATCACTCAAATA
 ATATCAACAAGAACCAGTCTGCGCTGTGCGAGTTCTATCGAGCGTCTGTCTTCTGGCTTGC
 GTATTAAACAGCGCGAAGGATGACGCCGAGGTGAGGCGATTGCTAACCGTTTTACTTCTA
 ACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAACGATGGTATTTCTGTTGCACAGA
 CCACTGAAGGCGCGCTGTCCGAAATCAACAACAACCTTACAGCGTATCCGTGAACTGACGG
 TTCAGGCTTCTACCGGGACTAACTCCGATTCTGGATCTGGACTCCATTCAGGACGAAATCA
 AATCCCGTCTGGACGAAATTGACCGCGTATCTGGCCAGACCCAGTTCAACGGCGTGAACG
 TACTGGCGAAAGACGGTTCAATGAAAATTCAGGTTGGTGCGAATGACGGCCAGACTATCA
 CGATTGATCTGAAGAAAATTGACTCTGATACGCTGGGGCTGAGTGGGTTTAAATGTGAATG
 GTAGCGGGGCTGTGGCTAATACTGCAGCGACTAAATCTGATTTGGCAGCAGCTCAACTCT
 TGGCTCCAGGTACTGCTGATGCTAATGGTACAGTTACCTATACTGTGGCGCAGGCTGA
 AAACATCTACAGCTGCAGATGTAATTGCGAGTTTGGCTAATAACGCAAAAGTTAATGCCA
 CAATTGCAAATGGTTTTGGATCGCCAACAGCTACAGATTATACATACAACAGCGCTACAG
 GCGATTTTACATATAGTGCAACTATTGCAGCTGGTACAAATTCTGGTGATAGTAACAGTG
 CTCAGTTACAATCCTTCTGACACCAAAAGCGGGCGATACTGCTAACTTAAACGTTAAAA
 TTGGTTCTACGTCAATTGACGTTGTATTGGCTAGCGACGGTAAATTTACCGCGAAAGATG
 GTTCAGAACTATTTATTGACGTAGATGGTAACCTCACTCAAAACAATGCTGGGACTGTCA
 AAGCAGCCACTCTTGATGCACTGACTAAAACTGGCATACAACAGGCACACCGAGTGCCG
 TATCTACGGTAATTACAACCTGAAGATGAAACAACCTTCACTCTGGCTGGCGGTACTGATG
 CTACTACTTCTGGTGCAATCACTGTAGCAAATGCAAGAATGAGTGCTGAGTCTCTTCAAT
 CGGCAACTAAGTCCACAGGATTCACAGTTGATGTTGGAGCTACTGGTACCAGCGCAGGCG
 ATATTAAAGTTGATAGTAAAGGTATAGTACAACAACACAGGTACAGGTTTTGAAGACG
 CTTACACCAAAGCTGATGGTTCAGTACTACCGATAATACAACCAATCTGTTTTTGCAA
 AAGACGGAAGTGTGACCAATGGTTCAGGTAAAGCAGTCTATGTTTCAGCGGATGGTAATT
 TTAATACTGACGCTGAACTAAAGCTGCAACCAACCGCCGATCCACTGAAAGCTCTGGACG
 AAGCGATCAGCTCCATCGACAAATTCCGTTCTTCCCTCGGTGCGGTGCAAAACCGTCTGG
 ATTCCGCGAGTCACCAACCTGAACAACACCACTACTAACCTGTCTGAAGCGCAGTCCCCTA
 TTCAGGACGCTGACTATGCGACCGAAGTGTCGAATATGTCGAAAGCGCAGATCATCCAGC
 AGGCCGGTAACTCCGTGCTGGCAAAAGCTAACCAGGTACCGCAGCAGGTTCTGTCTCTGC TGCAGGGTTAA

Figure 32

AACAAAAACCAGTCTGCGCTGTCGACTTCTATCGAGCGCCTCTCTT
CTGGTCTGCGCATTAAACAGCGCTAAAGATGACGCTGCGGGCCAGGCGATTGCTAACCGCT
TCACTTCTAACATCAAAGGTCTGACTCAGGCCGCACGTAACGCCAACGACGGTATCTCTC
TGGCGCAGACCACTGAAGGCGCACTGTCTGAAATCAACAACAACCTTGCAGCGTGTTCGTG
AGCTGACCGTTCAGGCCACTACCGGTACTAACTCTGATTCTGACCTGTCTTCAATCCAGG
ACGAAATCAAATCCCGTCTCGATGAAATTGACCGGTATCCGGTCAGACTCAGTTCAACG
GCGTGAACGTACTGGCAAAAGATAACACCATGAAGATTAGGTTGGTGCGAACGATGGTC
AGACTATATCCATCGACCTGCAAAAAATCGACTCTTCTACTCTTGGTTTGAACGGTTTCT
CCGTTTCTAAAAATGCTCTCGAAACTAGCGAAGCGATCACTCAGTTGCCGAACGGTGCGA
ATGCACCAATCGCTGTGAAGATGGATGCGTCTGTTCTGACCGATCTTAACATTACTGATG
CTTCCGCTGTTTCGCTGCACAACGTAACATAAGGTGGTGTGCAACGTCTACTTATGTTG
TTCAGTATGGCGATAAGAGCTATGCAGCATCTGTTGATGCGGGAGGTACAGTAAACTGA
ATAAAGCCGACGTAACATATAACGACGCAGCAAATGGTGTTACGAATGCCACCCAGATTG
GTAGTCTGGTTTCAGGTTGGTGCTGATGCAACAATGATGCAGTTGGTTTTGTTACCGTGC
AGGGGAAAAACTATGTTGCTAATGACTCATTAGTCAATGCTAATGGCGCTGCTGGCGCTG
CAGCAACTAGAGTTACAATTGATGGTGATGGTAGCCTTGGAGCTAACCAGGCTAAAATTG
AACTTAGCCAAAATGGTGCTACTGTGCAACATCAGAGTTCGCTGGTGCTTCAACCAACG
ATCCACTGACTCTGCTGGACAAAGCTATCGCATCTGTTGATAAATTCCGTTCTTCTTTGG
GGGCGGTACAGAACCGTCTGAGCTCCGCTGTAACCAACCTGAACAACACCCTACCAACC
TGTCTGAAGCGCAGTCCCGTATTAGGACGCCGACTATGCGACCGAAGTGTCCAACATGT
CGAAAGCGCAGATCATCCAGCAGGCAGGTAACCTCCGTGCTGTCCAAA

Figure 33

ATGGCACAAGTCATTAATACCAACAGCCTCTCGCTGATCACTCAAAATAATATCAACAAGA
 ACCAGTCTGCGCTGTCGAGTTCTATCGAGCGTCTGTCTTCTGGCTTGCGTATTAACAGCG
 CGAAGGATGACGCCGAGGTGAGGCGATTGCTAACCGTTTTACTTCTAACATTAAAGGCC
 TGACTCAGGCTGCACGTAACGCCAACGACGGTATTTCTGTTGCACAGACCACTGAAGGCG
 CGCTGTCCGAAATCAACAACAACTTACAGCGTATTCGTGAACTGACGGTTCAGGCGACGA
 CCGGAACTAACCTCCACCTCTGACCTGGACTCCATTGAGGACGAAATCAAATCCCGTCTTG
 ATGAAATTGACCGGTATCCGGCCAAACCCAGTTCAACGGCGTGAACGTACTGTCAAAAG
 ATGGCTCGATGAAAATTGAGTTCGGCGCAAATGATGGTGAAACCATCACGATTGATCTGA
 AAAAGATCGACTCTTCTACATTGAAGCTGACCAGCTTCAATGTTAACGGTAAAGGCGCTG
 TTGATAATGCTAAAGCCACTGAAGCAGATCTGACCGCTGCGGGCTTCTCCCAAGGTGCAG
 TCGTCACTGGCAACAGCACCTGGACTAAATCTACTGTTACTACCTTTAATGCAGCAACAG
 CTACCGACGTGCTGGCAAGCGTTAGCGGCGGCAGCACTATTAGCGGTTATACCGGTACAA
 ACAATGGATTAGGCGTAGCGGCTTCTACTGCATATACCTACAACGCAACCAGCAAGTCTT
 ATTCATTGACGCAACCGCACTTACCAATGGCGATGGTACTGGGGCCACCACTAAAGTTG
 CTGATGTGCTGAAAAGCCTATGCAGCAAACGGTGATAATACGGCTCAGATCTCCATCGGCG
 GAAGCGCTCAGGACGTTAAAATTGCCAGCGATGGCACCCTGACTGACGTCAATGGTGATG
 CTTTATATATTGGTTCTGACGGCAACCTGACTAAAAACCAGGCCGGCGGTCCAGATGCGG
 CAACGTTGGACGGTATTTTCAACGGTGCGAATGGTAATGCAGCAGTTGATGCGAAGATTA
 CATTCCGGCAGCGCATGACCGTTGATTTCACCCAGGCTAGCAAAAAAGTGGATATTAAGG
 GCGCAACGGTATCCGCCGAAGATATGGACACTGCGTTAACTGGGCAGGCTTATACCGTAG
 CTAACGGCGCACAGTCTTTTGACGTTGCCGCTGGTGGGGCAGTAACCGCTACTACAGGTG
 GCGCTACCGTAAATATTGGTGCTGATGGTGAACTGACGACTGCGACCAACAAGACTGTCA
 CAGAAACTTATCACGAATTTGCTAACGGCAATATTCTGGATGATGACGGCGCGGCTCTGT
 ACAAAGCGGCTGACGGTTCTCTGACCACTGAAGCTACTGGTAAATCCGAAGTGACCACGG
 ATCCGCTGAAAGCGCTGGACGATGCTATCGCATCCGTAGACAAATCCGCTCCTCCCTCG
 GTGCGGTGCAGAACCGTCTGGATTCCGCAGTCACCAACCTGAACAACCACTACCAACC
 TGTCTGAAGCGCAGTCCCGCATTCAGGACGCCGACTATGCGACCGAAGTGTCCAATATGT
 CGAAAGCGCAGATCATCCAGCAGGCCGGTAACTCCGTGCTGGCAAAAGCCAACCAGGTAC
 CGCAGCAGGTTCTGTCTCTGCTGCAGGGTTAA

Figure 34

ATGGCACAAGTCATTAATACCAACAGCCTCTCGCTGATCAC
 TCAAAATAATATCAACAAGAACCAGTCTGCGCTGTCGAGTTCTATCGAGCGTCTGTCTTC
 TGGCTTGCGTATTAAACAGCGCTAAGGATGACGCCGCGGGTCAGGCGATTGCTAACCGTTT
 TACTTCTAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAACGACGGTATTTCTGT
 TGCAGACCACTGAAGGCGCGCTGTCCGAAATCAACAACAATTACAGCGTATCCGTGA
 ACTGACGGTTTCAGGCTTCTACCGGGACTAACTCCGATTCGGATCTGGACTCCATTTCAGGA
 CGAAATCAAATCCCGTCTGGACGAAATTGACCGCGTATCTGGCCAGACCCAGTTCAACGG
 CGTGAACTGACTGGCGAAAGACGGTTCAATGAAAATTCAGGTTGGTGCGAATGACGGCCA
 GACTATCACTATTGATCTGAAGAAAATTGACTCAGATACGCTGGGGCTGAGTGGGTTTAA
 TGTGAATGGTGGCGGGCTGTTGCTAATACTGCAGCGACTAAAGATGATTTGGTCGCTGC
 ATCAGTTTCAGCTGCGGTAGGTAATGAATACACTGTCTCTGCTGGCCTGTGCGAAATCAAC
 TGCTGCTGATGTTATTGCTAGTCTCACAGATGGTGGCAGTAAGTGGCTGGTGTAAAG
 CAATGGTTTTGCTGCAGGGGCAACTGGAGATGCTTATAAATTCAATCAAGCAAACAACAC
 TTTTACTTACAATACCACCTCAACAGCGGCAGAACTCCAATCTTACCTCACGCCCTAAGGC
 GGGGATACCGCAACTTTCTCCGTTGAAATTGGTGGCACCAAGCAGGATGTTGTTCTGGC
 TAGTGATGGCAAAATCACAGCAAAAGACGGGTCTAACTTTATATTGACACCACAGGGAA
 TTTAACCCTAAACGGTGGAGGTACTTTAGAAGAAGCTACCCCTCAATGGCTTAGCTTTCAA
 CCACTCTGGTCCAGCCGCTGCTGTACAATCTACTATTACTACTGCGGATGGAACCTTCAAT
 AGTTCTAGCAGGTTCTGGCGACTTTGGAACAACAAAACTGCTGGGGCTATTAAATGTCAC
 AGGAGCAGTGATCAGTGCTGATGCACTTCTTTCCGCCAGTAAAGCGACTGGGTTTACTTC
 TGGCACTTATACCGTAGGTACAGATGGAGTTGTTAAATCTGGTGGCAATGACGTTTATAA
 CAAAGCTGACGGGACGGGATTAAGTACTGACAATACCACAAAATATTATTTACAAGATGA
 CGGGTCTGTAACATAATGGTTCTGGTAAAGCTGTGTATGCTGATGCAACAGGAAAATAAC
 TACTGACGCTGAACTAAAGCCGAAACCACCGCCGATCCCTGAAAGCTCTGGACGAAGC
 GATCAGCTCCATCGACAAATCCGTTCTTCCCTCGGTGCGGTGCAAAACCGTCTGGATTTC
 CGCGGTACCAACCTGAACAACCACTACCAACCTGTCCGAAGCGCAGTCCCGTATTCA
 GGACGCCGACTATGCGACCGAAGTGTCCAACATGTGAAAGCGCAGATCATCCAGCAGGC
 CGGTAATCCGTGCTGGCAAAAGCTAACCAGGTACCGCAGCAGGTTCTGTCTCTGCTGCA GGGTTAA

Figure 35

ATGGCACAAGTCATTAATACCAACAGCCTCTCGCTGATCAC
 TCAAAATAATATCAACAAGAACCAGTCTGCGCTGTCGAGTTCTATCGAGCGTCTGTCTTC
 TGGCTTGCGTATTAACAGCGCGAAGGATGACGCCGCGGGTCAGGCGATTGCTAACCGTTT
 TACTTCTAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAACGACGGTATTTCCGT
 TGCGCAGACCACCGAAGGCGCGTGTCCGAAATCAACAACAATTACAGCGTATCCGTGA
 ACTGACGGTTTCAGGCCACTACCGGTACTAACTCCGATTCTGACCTGGACTCCATCCAGGA
 CGAAATCAAATCTCGTCTTGATGAAATTGACCGCGTATCTGGTCAGACCCAGTTCAATGG
 CGTGAATGTGTTGTCCAAAGACGGTTCAATGAAAATTGAGGTGGGCGCAAATGATGGTGA
 AACCATCAGATTGACCTGAAAAAATCGACTCTTCTACACTGAAGCTGACCAGCTTCAA
 CGTCAACGGTAAAGGCGCTGTTGATAATGCAAAAGCCACTGAAGCAGATCTGACCGCTGC
 GGGCTTCTCCCAAAGTGCAAGTTGTCAGTGGCAATAGCACCTGGACTAAATCTACTGTTAC
 TACCTTTAATGCAGCAACAGCTACCGATGTGCTGGCTAGCGTTAGTGGCGGCAGCACTAT
 TAGCGGTTATGCTGGCACAACAATGGGTTAGGCGTAGCGGCTTCTACTGCATATACCTA
 CAACGCAACCAGCAAGTCTTATTCATTTGACGCAACCGCACTTACTAATGGTGATGGTAC
 TGCGGGCTCAACTAAAGTTGCTGATGTTCTGAAAGCCTATGCAGCAAACGGCGATAACAC
 GGCTCAGATCTCCATCGGTGGTAGCGCTCAGGAAGTTAAAATTGCCAGCGATGGTACCCT
 GACGGATACTAATGGCGATGCTTTATACATTGGTGCTGACGGTAACCTGACGAAAAACCA
 GGCCGGCGGGCCAGCCGCGGCAACGTTGGACGGTATTTTCAACGGTGCGAATGGTCATGA
 TGCAGTTGATGCGAAGATTACCTTCGGCAGCGGCATGACCGTTGACTTCACCCAGGTTAG
 CAACAATGTGGATATTAAGGGCGCGACGGTATCCGCCGAAGATATGAACACTGCGTTAAC
 CGGTGAGGCTTATACCGTAGCTAACGGCGCACAGTCTTATGACGTTGCCGCTGATGGTGC
 AGTAACTGCTACTACAGGTGGAGCGACCGTAAATATTGGTGCTGAGGGTGAAC TGACGAC
 TGCGGCCAACAAAGACTGTACAGAAACTTATCACGAATTTGCTAACGGCAATATTCTGGA
 TGATGACGGCGCGGCTCTGTATAAAGCGGCTGACGGCTCTTGACCACTGAAGCTACAGG
 TAAATCTGAAGCGACCACGGATCCGCTGAAAGCGCTGGACGATGCTATCGCATCCGTAGA
 CAAATTCGGTTCTTCCCTGGGTGCCGTGCAGAACCGTCTGGATTCCGCAGTCACCAACCT
 GAACAACACCACTACCAACCTGTCCGAAGCGCAGTCCCCTATTGAGGACGCCGACTATGC
 GACCGAAGTGTCACCATGTGCGAAGCGCAGATTATTGAGCAGGCAGGTAACTCCGTGCT
 GGCAAAAGCTAACCGGTACCGCAGCAGGTTCTGTCTCTGCTGCAGGGTTAA

Figure 36

AACAAAAACCAGTCTGCGCTGTCGACTTCTAT
CGAGCGCCTCTCTTCTGGTCTGCGCATTAAACAGCGCTAAAGATGACGCTGCGGGCCAGGC
GATTGCTAACCGCTTCACTTCTAACATCAAAGGTCTGACTCAGGCCGCACGTAACGCCAA
CGACGGTATCTCTTGGCGCAGACCACTGAAGGCGCACTGTCTGAAATCAACAACAACCTT
GCAGCGTGTGCGTGAGTTGACTGTTCAAGCGACGACCGGGACTAACTCTGATTCTGACCT
GTCTTCTATTCAAGACGAAATCAAATCCCGTCTGGATGAAATTGACCGTGTTCGGTCA
GACCCAGTTCAACGGCGTGACGTGCTGGCTAAAAACGGTCTATGGCGATTCAAGTTGG
CGCGAATGATGGGCAGACCATCAACATCGACCTGCAGAAAATCGACTCTTCTACTCTGGG
CCTGGGCGGCTTCTCCGTATCTAACAATGCACTGAACTGAGCGATTCTATCACTCAGGT
TGGTGCGAGTGGTTCAGTGGCAGATGTGAACTGAGCTCTGTTGCCTCGGCTCTGGGTGT
AGACGCAAGCACTCTGACTCTGCACAACGTACAGACCCAGCTGGCGCAGCAACAGCTAA
CTATGTTGTCTCTTCTGGTCTGACAACTACTCAGTATCTGTTGAAGATAGCTCCGGTAC
AGTTACGCTGAACACCACTGATATAGGTTATACCGATACCGCTAATGGCGTTACTACCGG
TTCCATGACTGGTAAGTACGTTAAAGTTGGAGCTGATGCATTGGGTGCTGCTGTAGGTTA
TGTCACCGTACAGGGACAAAACCTTCAAAGCTGATGCTGGCGCGTGGTTAACTCCAAGAA
TGCTGCTGGTAGTCAGAATGTTACTTCTGCAATTGGCGATATTGCTAATAAAGCGAATGC
TAACATTTACACTGGAACCTCTTCTGCAGATCCACTGGCTCTGCTGGACAAAGCTATCGC
ATCTGTTGATAAATTCGTTCTTCTCTAGGGGCGGTGCAGAACCGTCTGAGCTCTGCTGT
AACCAACCTGAACAACACCACTACCAACCTGTCCGAAGCGCAGTCCCGTATTCAGGACGC
CGACTATGCGACCGAAGTGTCCAACATGTCGAAAGCGCAGATCATCCAGCAGGCGGGTAA
CTCCGTGCTGTCTAAA

Figure 37

ATGGCACAAAGTCATTAATACCAACAGCCTCTCGCTGATCA
 CTCAAAATAATATCAACAAGAACCAGTCTGCGCTGTCGAGTTCTATCGAGCGTCTGTCTT
 CTGGCTTGCGTATTAAACAGCGCGAAGGATGACGCCGCCGGTCAGGCGATTGCTAACCGTT
 TTACTTCTAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAATGACGGTATTTCTG
 TTGCACAGACCACTGAAGGCGCGCTGTCCGAAATCAACAACACTTACAGCGTATTCGTG
 AACTGACGGTTCAGGCTTCTACCGGGACTAACTCTGATTCTGGATCTGGACTCCATTTCAGG
 ACGAAATCAAATCCCGTCTCGACGAAATTGACCGCGTATCCGGTCAGACCCAGTTCAACG
 GCGTGAACGTACTGGCAAAAGACGGTTCGATGAAAATTCAGGTTGGTGCGAACGACGGCC
 AGACTATCACTATTGATCTGAAGAAAATTGACTCTGATACGCTGGGGCTGAGTGGGTTTA
 ACGTAAATGGTAGCGCAGATAAGGCAAGTGTGCGGCGACAGCTGACGGAATGGTTAAAG
 ACGGATATATCAAAGGGTTAACTTCATCTGACGGCAGCACTGCATATACTAAACTACAG
 CAAATACTGCAGCAAAAGGATCTGATATTCTTGCGGCGCTTAAGACTGGCGATAAAATTA
 CCGCAACAGGTGCAAAATAGCCTTGCTGATAATGCGACATCGACAACCTTATACTTATAATG
 CAACCAGCAATACCTTCTCCTATACGGCTGACGGTGTAAACCAAACGAATGCTGCAGCAA
 ATCTCATACCTGCAGCAGGGAAAACGACAGCTGCATCAGTTACTATTGGTGGGACAGCAC
 AGAATGTAAATATTGATGATTGCGGCAATATTACTTCAAGTGATGGCGATCAACTTTATC
 TGGATTCAACAGGTAACCTGACTAAAACAGGCCGGCAACCCGAAAAAGCAACCGTTT
 CTGGGCTTCTCGGAAATACGGATGCGAAAGGTACTGCTGTTAAACAACCATCAAGACAG
 AGGCTGGTGTAACAGTTACAGCTGAAGGTAATACAGGTACTGTAAAAATTGAAGGTGCTA
 CTGTTTCAGCATCTGCATTTACGGGCATTGCATATTCGCCAACACCGGTGGGAATACTT
 ATGCTGTTGCCGCAATAATACTACAAATGGTTTCCTGGCGGGGGATGACTTAACCCAGG
 ATGCTCAAACCTGTTTCAACCTACTACTCGCAAGCCGATGGCACGGTCACGAATAGCGCAG
 GCAAAGAAATCTATAAAGACGCTGATGGTGTCTACAGCACAGAGAATAAAACATCGAAGA
 CGTCCGATCCATTGGCTGCGCTTGACGACGCAATCAGCTCCATCGACAAATTCGGTTCAT
 CCTTGGGTGCTATCCAGAACCGTCTGGATTCCGCGGTACCAACCTGAACAACACCACTA
 CCAACCTGTCCGAAGCGCAGTCCCGTATTCAGGACGCCGACTATGCGACCGAAGTGTCCA
 ACATGTCGAAAGCGCAGATCATCCAGCAGGCCGGTAACTCCGTGCTGGCAAAAGCTAACC
 AGGTACCGCAGCAGGTTCTGTCTCTGCTGCAGGGCTAA

Figure 38

AACAAATCTCAGTCTTCTCTGAGCTCCGCCATTGAACGTCTCTCTTCTGGCCTGCGTA
TTAACAGTGCTAAAGATGACGCAGCAGGTCAGGCGATTGCTAACCGTTTTACAGCAAATA
TTAAAGGTCTGACTCAGGCTTCCCGTAACGCGAATGATGGTATTTCTGTTGCGCAGACCA
CTGAAGGTGCGCTGAATGAAATTAACAACAACCTGCAGCGTGACGTGAACTGACTGTTC
AGGCAACTAACGGTACTAACTCTGACAGCGATCTTTCTTCTATCCAGGCTGAAATTACTC
AACGTCTGGAAGAAATTGACCGTGTATCTGAGCAAACCTCAGTTTAACGGCGTGAAAGTCC
TTGCTGAAAATAATGAAATGAAAATTACGGTTGGTGCTAATGATGGTGAAACCATCACTA
TCAATCTGGCAAAAATTGATGCGAAAACCTCTCGGCCTGGACGGTTTTAATATCGATGGCG
CGCAGAAAGCAACTGGCAGTGACCTGATTTCTAAATTTAAAGCGACAGGTACTGATAACT
ATGATGTTGGCGGTGATGCTTATACTGTTAACGTAGATAGCGGAGCTGGGTAAAGACTCC
AACTTATTGATAGTGTTTTATGTTTCAAGATAATGCCCGATGACTTTGTGATGCAGCTCCAC
CGATTTTGAGAACGACAGCGACTTCCGTCCAGCCGTGCCAGGTGCTGCCTCAGATTCAG
GTTATGCCGCTCAATTGCTGCGTATATCGCTTGCTGATTACGTGCAGCTTTCCCTTCAG
GCGGGATTATACAGCGGCCAGCCATCCGTCATCCATATCACCACGTCAAAGGGTGACAG
CAGGCTCATAAGACGCCCCAGCGTCGCCATAGTGCGTTTACCAGAATACGTGCGCAACAAC
CGTCTTCCGGAGCCTGTATACGCGTAAAACAGCCAGCGCTGGCGCGATTTAGCCCCGAC
ATAGTCCCACTGTTTCGTCCATTTCCGCGCAGACGATGACGTCACTGCCCGGTGTATGCG
CGAGGTTACCGACTGCGGCCTGAGTTTTTTAAGTGACGTAAAATCGTGTTGAGGCCAACG
CCCATAATGCGGCGAGTTGCCCGGCATCCAACGCCATTCATGGCCATATCAATGATTTTC
TGGTGCGTACCGGGTTGAGAAGCGGTGTAAGTGAAGTGCAGTTGCCATGTTTTACGGCAG
TGAGAGCAGAGATAGCGCTGATGTCCGGCGGTGCTTTTGCCGTTACGCACCACCCCGTCA
GTAGCTGAACAGGAGGGACAGCTGATAGAAAACAGAAGCCACTGGAGCACCTCAAAAACAC
CATCATACACTAAATCAGTAAGTTGGCAGCATTACCGCGGAGCTGTTAAAGATACTACAG
GGAATGATATTTTGTAGTGACGAGATGGTTCACTGACAACATAATCTGACACAAACA
TAGCTGGTACAGGGATTGATGCTACAGCACTCGCAGCAGCGGCTAAGAATAAAGCACAGA
ATGATAAATTCACGTTTAAATGGAGTTGAATTCACAACAACAACCTGCAGCGGATGGCAATG
GGAATGGTGTATATTCTGCAGAAATTGATGGTAAGTCAGTGACATTTACTGTGACAGATG
CTGACAAAAAGCTTCTTTGATTACGAGTGAGACAGTTTACAAAAATAGCGCTGGCCTTT
ATACGACAACCAAAGTTGATAACAAGGCTGCCACACTTTCCGATCTTGATCTCAATGCAG
CTAAGAAAACAGGAAGCACGTTAGTTGTTAACGGTGCAACTTACGATGTTAGTGCAGATG
GTAAAACGATAACGGAGACTGCTTCTGGTAACAATAAAGTCATGTATCTGAGCAAATCAG
AAGGTGGTAGCCGATTCTGGTAAACGAAGATGCAGCAAATCGTTGCAATCTACCACCA
ACCCGCTCGAACTATCGACAAAGCATTGGCTAAAGTTGACAATCTGCGTTCTGACCTCG
GTGCAGTACAAAACCGTTTCGACTCTGCTATCACCACCTTGGCAACACCGTAAACAACC
TGTCTTCTGCCCGTAGCCGTATCGAAGATGCTGACTACGCGACCGAAGTGTCTAACATGT
CTCGTGCGCAGATCCTGCAACAAGCGGTACCTCTGTCTGGCGCAG

Figure 39

AACAAGAACCAGTCTGCGCTGTCGAGTTCTATCGAGCGTCTGT
CTTCTGGCTTGCGTATTAACAGCGCGAAGGATGACGCCGAGGTCAGGCGATTGCTAACC
GTTTTACTTCTAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAACGACGGTATTT
CTGTTGCGCAGACCACCGAAGGCGCGCTGTCCGAAATCAACAACAACCTTACAGCGTGTGC
GTGAACTGACCGTTTCAGGCAACCACCGGTACCAACTCCCAGTCTGACCTGGACTCTATCC
AGGACGAAATTAATCCCGTCTGGACGAAATTGACCGGTATCCGGTCAGACCCAGTTCA
ACGGCGTGAACGTACTGGCAAAAGACCGTTCCATGAAAATTCAGGTTGGCGCGAACGATG
GCCAGACCATCACTATCGACCTGAAGAAGATTGACTCTTCTACGCTGAAACTGACTGGTT
TTAACGTGAATGGCAAAGCAGCGGTTGATAATGCTAAAGCGACGGATGCAAATCTGACTA
CCGCCGGTTTTACACAAGGCGTTGTGGATTCAAATGGTAATAGTACTTGGACTAAATCAA
CTACGACTAATTTTCGATGCGCAACTGCAGTAAACGTACTAGCAGCAGTTAAAGATGGCA
GCACAATCAATTACACCGGTACTGGTAATGGTTTAGGGATTGCTGCAACAAGTGCTTATA
CATATCACGATAGCACTAAATCCTATACCTTTGATTCTACGGGGGCTGCAGTAGCTGGTG
CCGCGTCCAGCCTGCAAGGTACTTTTGGTACAGATACGAATACTGCAAAAATCACCATCG
ATGGTTCTGCTCAAGAAGTAAACATCGCTAAAGATGGGAAAATTACTGATACTGATGGTA
AAGCTTTATATATCGATTCCACTGGTAATTTGACTAAGAACGGCTCTGATACTTTAACTC
AGGCAACATTGAATGATGTCCTTACTGGTGCTAATTCAGTTGATGATACAAGGATTGACT
TCGATAGCGGCATGTCTGTACCCCTTGATAAAGTGAACAGCACTGTAGATATCACTGGCG
CATCTATTTTCAGCCGCTGCAATGACTAATGAGTTGACAGGTAAGGCCTATACCGTAGTAA
ATGGTGCGAATCTTACGCTGTAGCTACTAATAACACAGTAAAAACGACTGCTGATGCTA
AAAATGTTTATGTTGATGCTAGTGGTAAATTAATACTGATGACAAAGCCACTGTTACAG
AACTTATCATGAATTTGCGAATGGCAATATCTATGATGATAAAGGCGCTGCTGTTTATG
CGCGGCGGATGGTTCTCTGACTACAGAACTACAAGTAAATCAGAAGCTACAGCTAACC
CGCTGGCCGCTCTGGACGACGCAATCAGCCAGATCGACAAATCCGTTTCATCCCTGGGTG
CTATCCAGAACCGTCTGGATTCCGCACTACCAACCTGAACAACCACTACCAATCTGT
CTGAAGCGCAGTCCCGTATTTCAGGACGCCGACTATGCGACCGAAGTGTCCAATATGTCGA
AAGCGCAGATCATCCAGCAGGCAGGCAACTCCGTGCTGGCAAAA

Figure 40

AACAAAAACCAGTCTGCGCTGTCGACTTCTATCGAGCGCCTCTC
TTCTGGTCTGCGCATTAAACAGCGCTAAAGATGACGCTGCGGGCCAGGCGATTGCTAACCG
CTTCACCTTCTAACATCAAAGGTCTGACTCAGGCCGCACGTAACGCCAACGACGGTATCTC
TCTGGCGCAGACCACTGAAGGCGCACTGTCTGAAATCAACAACAACCTTGCAGCGTGTTCG
TGAAGTGACCGTTCAGGCCACTACCGGTACTAACTCTGATTCTGACCTGTCTTCAATCCA
GGACGAAATCAAATCCCGTCTCGATGAAATTGACCGCGTATCCGGTCAGACTCAGTTCAA
CGGCGTGAACGTACTGGCAAAAGATGGCTCGATGAAATTTCAGGTCGGTGCAAATGATGG
TCAGACAATCAGCATTGATTTGCAGAAGATTGATTCTTCTACTTTAGGGTTAAATGGTTT
TTCTGTTTCCAAAAATGCAGTATCTGTTGGTGATGCTATTACTCAATTGCCTGGCGAGAC
GGCAGCCGATGCACCAGTAACCATCAAGTTTGATGATTCAAGTAAAACTGATTTAAAACT
GACCGATGCTTCAGGGTTAAGTCTGCATAACCTCAAAGATGAAATGGTAATTTAACTAA
CCAGTATGTTGTACAGAATGGCGGAAATCTTACGCTGCTACAGTCGCTGCCAATGGTAA
TGTTACGCTGAACAAAGCAAATGTAACTACAGCGATGTCGCAACGGTATTGATACCGC
AACGCAGTCAGGCCAGTTAGTTCAGGTTGGTGCAGATTCTACCGGTACGCCAAAAGCATT
CGTGTCTGTCCAAGGTAAAAGCTTTGGCATTGATGACGCCCGCTTGAAGAATAACACTGG
TGATGCTACCGCTACTCAACCGGGAACATCTGGGACAACAGTTGTGCGAGCGTCAATTCA
TCTGAGTACGGGCAAAAACTCTGTAGACGCTGATGTAACGGCTTCCACTGAATTCACAGG
TGCTTCAACCAACGATCCACTGACTCTGCTGGACAAAGCTATCGCATCTGTTGATAAATT
CCGTTCTTCTTTGGGGCGGTACAGAACCGTCTGAGCTCCGCTGTAACCAACCTGAACAA
CACCACCACCAACCTGTCTGAAGCGCAGTCCCGTATTCAGGACGCCGACTATGCGACCGA
AGTGTCCAACATGTCGAAAGCGCAGATTATCCAGCAGGCAGGTAACCTCCGTGCTGTCCAA A

Figure 41

AACAAAAACCAGTCTGCGCTGTCGACTTCTATCGAGCGCCTCTCTTCTGGTC
TGCGCATTAAACAGCGCTAAAGATGACGCTGCGGGCCAGGCGATTGCTAACCGCTTCACTT
CTAACATCAAAGGTCTGACTCAGGCTGCACGTAACGCCAATGACGGTATTTCTCTAGCAC
AGACAGCGGAAGGCGCGCTGTCAGAGATTAACAACAACCTGCAGCGTGTGCGTGAGTTGA
CCGTGCAGGCAACCACTGGTACCAACTCTGATTCCGATCTCTTCTATTTCAGGATGAAA
TTAAATCTCGTCTGGATGAAATTGACCGCGTCTCTGGTCAGACCCAGTTTAAACGGCGTGA
ACGTACTGGCTAAAAACGGTTCTATGGCAATTCAGGTTGGCGCGAACGATGGCCAGACTA
TCTCTATCGACCTGCAGAAAATAGACTCTTCTACTCTGGGTCTGAGCGGCTTCTCTGTTT
CTCAGAACTCCCTGAAACTGAGCGATTCTATCACTACGATCGGCAATACTACTGCTGCAT
CGAAGAACGTGGACCTGAGCGCAGTAGCAACTAACTGGGCGTGAATGCAAGCACCCCTGA
GCCTGCACGAAGTTCAGGACTCTGCTGGTGACGGTACTGGTACCTTCGTTGTTTCTTCTG
GCAGCGACAACCTATGCTGTGTCTGTAGACGCGCCTCTGGTGCAGTTAACCTGAACACCA
CTGACGTCACCTATGATGACGCTACTAATGGTGTTACTGGCGCGACTCAGAACGGTCAGC
TGATCAAAGTAACTTCTGACGCCAACGGTGACGCTGTTGGTTACGTAACCATTTCAGGGTA
AAAACTATCAGGCTGGTGCGACCGGTGTTGACGTTCTGGCGAACAGCGGTGTTGCAGCTC
CACTACAGCTGTTGATACCGGTACTCTGCAACTGAGCGGTACTGGTGCAACTACTGAGC
TGAAAGGTACTGCAACTCAGAACCCACTGGCACTATTGGACAAAGCTATCGCTTCTGTTG
ATAAATTCGGTTCTTCTCTGGGTGCGGTACAGAATCGTCTGAGCTCTGCTGTAACCAACC
TGAATAACACCACCCTAACCTGTCTGAAGCGCAGTCCCGTATTTCAGGATGCCGACTATG
CGACCGAAGTGTCAAATATGTCTAAAGCGCAGATCGTTCAGCAGGCCGGTAAC

Figure 42

AACAAATCTCAGTCTTCTCTTAGCTCTGCTATTGAGCGTCTGTCTTCT
 GGTCTGCGTATTAAACAGCGCAAAAGACGATGCAGCAGGTCAGGCGATTGCTAACCGTTTT
 ACGGCAAAATATTAAAGGTCTGACCCAGGCTTCCCGTAACGCAAATGATGGTATTCTGT
 GCGCAGACCACTGAAGGTGCGCTGAATGAAATTAACAACAACCTGCAGCGTATTCTGTGAA
 CTTTCTGTTCAGGCAACTAACGGTACTAACTCTGACAGCGATCTTCTTCTATCCAGGCT
 GAAATTACTCAACGTCTGGAAGAAATTGACCGTGTATCTGAGCAAACCTCAGTTTAACGGC
 GTGAAAGTCCTTGCTGAAAATAATGAAATGAAAATTCAGGTTGGTGCTAATGATGGTGAA
 ACCATCACTATCAATCTGGCAAAAATTGATGCGAAAACCTCTCGGCCTGGACGGTTTTAAT
 ATCGATGGCGCGCAGAAAGCAACAGGCAGTGACCTGATTCTAAATTTAAAGCGACAGGT
 ACTGATAATTATGATGTTGGCGGTAAAACCTTATACCGTGAATGTGGAGAGCGGCGCGGTT
 AAGAATGATGCTAATAAGATGTTTTTGTAAAGCGCAGCTGATGGATCGCTGACGACCAGT
 AGTGATACTAAAGTATCCGGTGAAAGTATTGATGCAACAGAAGTAGCGAAACTTGCAATA
 AAATTAGCTGACAAAGGCTCCATTGAATACAAGGGCATTACATTTACTAACAACACTGGC
 GCAGAGCTTGATGCTAATGGTAAAGGTGTTTTGACCGCAAATATTGATGGTCAAGATGTT
 CAATTTACTATTGACAGTAATGCACCCACGGGTGCCGCGCAACAATAACTACAGACACA
 GCTGTTTACAAAAACAGTGCGGGCCAGTTCACCACTACAAAAGTGGAATAAAGCCGCA
 AACTCTCTGATCTGGATCTTAATGCAGCCAAGAAAACAGGTAGCACTTTAGTTGTAAAT
 GCGGCCACCTACAATGTCAGCGCAGATGGTAAAACGGTAACTGATACTACTCCTGGTGCC
 CCTAAAGTGATGTATCTGAGCAAATCAGAAGGTGGTAGCCCGATTCTGGTAAACGAAGAT
 GCAGCAAAATCGTTGCAATCTACCACCAACCCGCTCGAAACTATCGACAAGGCATTGGCT
 AAAGTTGACAATCTGCGTTCTGACCTCGGTGCAGTACAAAACCGTTTCGACTCTGCCATC
 ACCAACCTTGGCAACACCGTAAACAACCTGTCTTCTGCCCGTAGCCGTATCGAAGATGCT
 GACTACGCGACCGAAGTGTCTAACATGTCTCGTGCGCAGATCCTGCAACAAGCGGGTACC
 TCTGTTCTGGCGCAG

Figure 43

ATGGCACAAGTCATTAATACCAACAGCCTCTCGCTGATCACT
 CAAAATAATATCAACAAGAACCAGTCTGCGCTGTCGAGTTCTATCGAGCGTCTGTCTTCT
 GGCTTGCGTATTAAACAGCGCGAAGGATGACGCCGAGGTGAGCGATTGCTAACCGTTTC
 ACCTCTAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAACGACGGTATTTCTGTT
 GCACAGACCACCGAAGGCGCGCTGTCCGAAATCAACAACAATTACAGCGTATCCGTGAA
 CTGACGGTTCAGGCTTCTACCGGGACTAACTCTGATTCTGGATCTGGACTCCATTGAGGAC
 GAAATCAAATCCCGTCTGGACGAAATTGACCGCGTATCCGGCCAGACCCAGTTCAACGGC
 GTGAACGTGCTGGCGAAAGACGGTTCAATGAAAATTGAGTTGGTGCGAATGACGGCCAG
 ACTATCACTATTGATCTGAAGAAAATTGACTCTGATACTCTGGGTTTGAGTGGATTAAAT
 GTGAATGGCAAAGGGGCTGTGGCTAACGCAAAGCGACCGAAGCAGATTAAACGGGGGCT
 GGTTTCTCTCAAGGAGCGGTGGATACAAACGGAAATAGTACTTGGACAAAATCAACCACC
 ACCAATTACTCAGCTGCAACAACCTGCTGACTTGTATCGACCATTAAGGATGGCTCTACT
 GTTACATATGCAGGGACAGACACCGGATTAGGGGTCGAGCAGCAGGAAATTATACTTAT
 GATGCGAACAGTAAATCTTATTCCTTCAATGCCAATGGTCTGACGGGCGCAAATACCGCA
 ACTGCACTCAAAGTTACTTGGGGACAGGTGCTAACACCGCTAAAATTTCTATCGGTGGT
 ACAGAGCAGGAAGTGAATATTGCCAAAGATGGCACTATTACAGATACGAATGGTGATGCG
 CTCTATCTGGATATTACCGGCAACCTGACTAAGAACTATGCGGGTTCACCACCTGCAGCA
 ACGCTGGATAACGTATTAGCTTCCGCAACTGTAAATGCCACTATCAAGTTTGATAGCGGT
 ATGACGGTTGATTACACTGCAGGTACTGGCGCGAATATTACAGGTGCATCCATTTCTGCA
 GATGACATGGCCGCAAACTGAGCGGAAAGGCGTACACTGTTGCCAATGGTGCTGAGTCT
 TATGACGTTGCTGCAGTTACGGGGGCTGTAACTACAGCAGGTAATTCACCTGTGTAT
 GCCGATGCAGACGGTAAATTAACGACGAGTGCCAGTAATACGGTTACTCAGACTTATCAC
 GAGTTTGCTAATGGTAAATTTATGATGACAAAGGCTCGTCACTGTATAAAGCTGCAGAT
 GGCTCTCTGACTTCTGAAGCTAAAGGGAAATCTGAAGCAACCGCCGATCCCCTGAAAGCT
 CTGGACGAAGCCATCAGCTCCATCGACAAATTCGCTCCTCCCTCGGTGCCGTTCAAAC
 CGTCTGGATTCTGCGGTGACCAACCTGAACAACACCCTACCAACCTGTCTGAAGCGCAG
 TCCCGTATTGAGGACCGGACTATGCGACCGAAGTGTCCTAATATGTCGAAAGCGCAGATC
 ATCCAGCAGGCCGGTAACTCCGTGTTGGCAAAGCTAACCAGGTACCGCAGCAGGTTCTG
 TCTCTGCTGCAGGGTTAA

Figure 44

GCGCTGTCGACTTCTATCGAGCGCCTCTCTTCTGGTTTGCGCATTAACAGCGCTA
 AAGATGACGCTGCGGGCCAGGCGATTGCTAACCGCTTCACTTCTAACATCAAAGGTCTGA
 CTCAGGCCGCACGTAACGCCAACGACGGTATCTCTCTGGCGCAGACCACTGAAGGCGCAC
 TGTCTGAAATCAACAACAACCTTGCAGCGTGTTTCGTGAACTGACCGTTCAGGCCACTACCG
 GTACTAACTCTGATTCTGACCTGTCTTCAATCCAGGACGAAATCAAATCCCGCTTGGCTG
 AAATCGATCGTGTCTCTGGTCAGACCCAGTTCAACGGCGTGAACGTGCTGGCTAAAAACG
 GTTCTCTGAATATTAGGTTGGCGCGAATGATGGGCAGACCATCTCTATCGATTTGCAGA
 AAATAGACTCTTCTGCCCCTTGGTTTAAGTGGTTTTAGTGTTGCCGGTGGGGCGCTAAAAT
 TAAGCGATACAGTGACGCAGGTGCGCGATGGTTTCAAGCGCGCCAGTTAAAGTGGATCTGG
 ATGCAGCAGCAACAGATATTGGTACTGCTTTGGGGCAAAGGTTAATGCAAGTTCTTTAA
 CGTTGCACAATATCTTAGACAAAGATGGTGCGGCAACTGAGAACTATGTTGTTAGCTATG
 GTAGTGATAATTACGCTGCATCTGTTGCAGATGACGGGACTGTAACCTTAATAAAACGG
 ATATTACTTATTCAGGCGGTGATATTACCGGCGCTACCAAAGATGATACGTTGATTAAAG
 TTGCTGCTAATTCTGACGGAGAGGCCGTTGGTTTCGCTACCGTTTCAAGGTAAGAATTATG
 AAATTACAGATGGTGTAACCAACAGTCCACTGCTGCACCAACCGATATTGCTCAGACCA
 TTGATCTGGATACGGCTGATGAATTTACTGGGGCTTCCACTGCTGATCCACTGGCACTTT
 TAGACAAAGCTATTGCACAGGTTGATACTTTCCGCTCCTCCCTCGGTGCCGTTCAAACCC
 GTCTGGATTCCGCAGTCACCAACCTGAACAACACTACTACCAACCTGTCTGAAGCGCAGT
 CCCGTATTCAAGACGCCGACTATGCGACCGAAGTGTTCCAATATGTGAAAGCGCAGATCA TCCAGCAGGCC

Figure 45

ATGGCACAAGTCATTAATACCAACAGCCTCTCGCTGATCACT
 CAAAATAATATCAACAAGAACCAGTCTGCGCTGTCGAGTTCTATCGAGCGTCTGTCTTCT
 GGCTTGCGTATTAACAGCGCGAAGGATGACGCAGCGGGTCAGGCGATTGCTAACCGTTTT
 ACTTCTAATATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAATGACGGTATTTCTCTG
 GCGCAGACCACTGAAGGCGCACTGTCTGAAATCAACAACAACCTGACAGCGTGTGCGTGAA
 CTGACCGTACAGGCGACAACCGGAACGAACTCCGAATCTGACCTGTCTCTATCCAGGAC
 GAAATCAAATCCCGTCTGGAAGAGATTGACCGCGTATCCGGCCAGACTCAGTTCAACGGC
 GTGAATGTGCTGGCAAAAGACGGCACCATGAAAATTAGGTAGGCGCGAACGATGGTCAG
 ACTATCTCTATCGATCTGAAAAAATCGACTCTTCAACCCTGGGCCTGACCGGTTTTGAT
 GTTTCGACGAAAGCGAATATTTCTACGACAGCAGTAACGGGGGCGGCAACGACCACTTAT
 GCTGATAGCGCCGTTGCAATTGATATCGGAACGGATATTAGCGGTATTGCTGCTGATGCT
 GCGTTAGGAACGATCAATTTGATAATACAACAGGCAAGTACTACGCACAGATTACCACT
 GCGGCCAATCCGGGCCTTGATGGTGCTTATGAAATCCATGTTAATGACGCGGATGGTCC
 TTCACTGTAGCAGCGAGTGATAAACAAGCGGGTGCTGCTCCGGGTACTGCTCTGACAAGC
 GGTAAAGTTCAGACTGCAACCACCACGCCAGGTACGGCTGTTGATGTCACTGCGGCTAAA
 ACTGCTCTGGCTGCAGCAGGTGCTGACACGAGTGGCCTGAACTGGTTCACTGTCCAAC
 ACGGATTCCGCAGGTAAAGTGACCAACGTGGGTACGGCTGCAGAATGACAGCGGCACT
 ATCTTTGCAACCGACTACGATGGCACCCTGTGACCACGCCGGGCGCAGAGACTGTGACT
 TACAAAGATGCTTCCGGTAACAGCACCACTGCGGGTGTCACTGGGTGGCTCTGATGGC
 AAAACCAATCTGGTTACCGCCGCTGACGGCAAAACGTACGGTGCGACTGCACTGAATGGT
 GCTGATCTGTCCGATCCTAATAACACCGTTAAATCTGTTGACAGACAACGCTAAACCGTTG
 GCTGCCCTGGATGATGCAATTGCGATGGTCGACAAATTCCGCTCCTCCCTCGGTGCGGTG
 CAAAACCGTCTGGATTCCGCAGTCACCAACCTGAACAACCACTACCAACCTGTCTGAA
 GCGCAGTCCCGTATTACAGACGCCGACTATGCGACCGAAGTGTCCAACATGTGAAAGCG
 CAGATTATCCAGCAGGCAGGTAACCTCCGTGCTGTCCAAAGCTAACAGGTTCCGCAGCAG
 GTTCTGTCTCTGCTGCAGGGTTAA

Figure 46

AACAAAAACCAGTCTGCGCTGTGCGACTTCTATCGAGCGCCTCTCTTCTGGT
CTGCGTATTAACAGCGCTAAAGATGACGCCGCGGGCCAGGCGATTGCTAACCGCTTACT
TCTAACATCAAAGGTCTGACTCAGGCCGCACGTAACGCCAACGACGGTATTTCTCTGGCG
CAGACGGCTGAAGGCGCGCTGTGAGAGATTAACAACAACCTTGACGCGTATTCGTGAACTG
ACCGTTCAGGCCTCTACCGGCACGAACTCTGATTCCGACCTGTCTTCTATTACAGGACGAA
ATCAAATCCCGTCTTGATGAAATTGACCGTGTATCTGGTCAGACCCAGTTCAACGGTGTG
AACGTGCTGTGAAAAACGATTCGATGAAGATTAGATTGGTGCCAATGATAACCAGACG
ATCAGCATTGGCTTGCAACAAATCGACAGTACCACTTTGAATCTGAAAGGATTTACCGTG
TCCGGCATGGCGGATTTAGCGCGGCGAACTGACGGCTGCTGATGGTACAGCAATTGCT
GCTGCGGATGTCAAGGATGCTGGGGGTAAACAAGTCAATTTACTGTCTTACACTGACACC
GCGTCTAACAGTACTAAATATGCGGTCGTTGATTCTGCAACCGGTAAATACATGGAAGCC
ACTGTAGCCATTACCGGTACGGCGGCGGCGTAAGTGTGGTGCAGCGGAAGTGGCGGGA
GCCGCTACAGCCGATCCGTTAAAGCACTGGATGCCGCAATCGCTAAAGTCGACAAATTC
CGCTCCTCCCTCGGTGCCGTTCAAAACCGTCTGGATTCTGCGGTACCAACCTGAACAAC
ACCACCACCAACCTGTCTGAAGCGCAGTCCCGTATTCAGGACGCCGACTATGCGACCGAA
GTGTCCAACATGTCGAAAGCGCAGATTATCCAGCAGGCCGGTAACTCCGTGCTGGCAAA

Figure 47

ATGGCACAAGTCATTAATACCAACAGCCTCTCGCTGATCACTC
 AAAATAATATCAACAAGAACCAGTCTGCGCTGTCGAGTTCTATCGAGCGTCTGTCTTCTG
 GCTTGCGTATTAAACAGCGCGAAGGATGACGCAGCGGGTCAGGCGATTGCTAACCGTTTTA
 CCTCTAACATTAAAGGTCTGACTCAGGCTGCACGTAACGCCAACGACGGTATTTCTGTTG
 CACAGACCACTGAAGGCGCGCTGTCCGAAATCAACAACAATTACAGCGTATCCGTGAAC
 TGACGGTTCAGGCTTCTACCGGGACTAACTCCGATTCCGATCTGGACTCCATTCAAGACG
 AAATCAAATCCCGTCTGGACGAAATTGACCGGTATCCGGTCAAACCCAGTTCAACGGTG
 TGAACGTACTGGCGAAAGACGGTTCGATGAAAATTCAAGTTGGTGCGAATGACGGCCAGA
 CTATCACGATTGATCTGAAGAAAATTGACTCAGATACGCTGGGGCTGAATGGTTTCAACG
 TTAATGGCAAAGGCACTATTGCGAACAAGCTGCTACAGTCAGCGATCTGACCGCTGCTG
 GTGCAACGGGAACAGGTCTTATGCTGTGACCACAAACAATACAGCACTCAGCGCTAGCG
 ATGCACTGTCTCGCCTGAAAACCGGAGATACAGTTACTACTACTGGCTCGAGTGCTGCGA
 TCTATACTTATGATGCGGCTAAAGGGAACCTTCACTCAAGCAACAGTTGCAGATGGCG
 ATGTTGTTAACTTTGCGAATACTCTGAAACCAAGCGGGTGGCACTACTGCATCAGGTGTTT
 ATACTCGTAGTACTGGTGATGTGAAGTTTGATGTAGATGCTAATGGCGATGTGACCATCG
 GTGGTAAAGCCGCTACCTGGACGCCACTGGTAACCTATCTACAAACAACCCCGGCATTG
 CATCTTCAGCGAAATTGTCCGATCTGTTTGCTAGCGGTAGTACCTTAGCGACAACCTGGTT
 CTATCCAGCTGTCTGGCACAACCTATAACTTTGGTGACGCGCAACTTCTGGCGTAACCT
 ACACCAAACTGTAAGCGCTGATACTGTACTGAGCACAGTGCAGAGTGCTGCAACGGCTA
 ACACAGCAGTTACTGGTGCGACAATTAAGTATAATACAGGTATTCAGTCTGCAACGGCGT
 CCTTCGGTGGTGTGAATACTAATGGTGCTGGTAATTCGAATGACACCTATACTGATGCAG
 ACAAGAGCTCACCACAACCGCATCTTACACTATCAACTACAACGTCGATAAGGATACCG
 GTACAGTAAGTGTAGCTTCAAATGGCGCAGGTGCAACTGGTAAATTGTCAGCTACTGTTG
 GGGCACAGGCTTATGTTAACTCTACAGGCAAACTGACCACTGAAACCACAGTGCAGGCA
 CTGCAACCAAAGATCCTCTGGCTGCCCTGGATGAAGCTATCAGCTCCATCGACAAATTCC
 GTTCATCCCTGGGTGCTATCCAGAACCGTCTGGATTCCGCGGTTACCAACCTGAACAACA
 CCACTACCAACCTGTCCGAAGCGCAGTCCCGTATTCAAGACGCCGACTATGCGACCGAAG
 TGTCCAACATGTGAAAAGCGCAGATTATCCAGCAGGCCGGTAACCTCCGTGCTGGCAAAAG
 CCAACCAGGTACCGCAGCAGGTTCTGTCTCTGCTGCAGGGTTAA

Figure 48

ATGGCACAAAGTCATTAATACCAACAGCCTCTCGCTGATCAC
 TCAAAATAATATCAACAAGAACCAGTCTGCGCTGTCGAGTTCTATCGAGCGTCTGTCTTC
 TGGCTTGGCTATTAACAGCGCGAAGGATGACGCCGAGGTGAGGCGATTGCTAACCGTTT
 TACTTCTAATATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAATGACGGTATTTCTGT
 TGCACAGACCACTGAAGGCGCGCTGTCCGAAATCAACAACAACCTTACAGCGTGTGCGTGA
 ACTGACCGTTTCAAGCGACCAACCGGTACCAACTCCCAGTCTGATCTGGACTCTATCCAGGA
 CGAAATCAAATCCCGTCTGGACGAAATTGACCGCGTATCCGGTCAGACTCAGTTCAACGG
 CGTGAACGTACTGGCAAAAGACGGTTCCATGAAAATTCAGGTTGGCGCGAATGATGGCCA
 GACCATCACTATCGACCTGAAGAAGATTGACTCTTCTACGTTGAAACTGACTGGTTTTAA
 CGTGAATGGTTCTGGTTCTGTGGCGAATACTGCGGCGACTAAAGACGAACTGGCTGCTGC
 TGCTGCGGCGGCGGGTACAACCTCCTGCTGTGCGTACTGACGGCGTGACCAAATATACCGT
 AGACGCAGGGCTTAACAAAGCCACAGCAGCAAACGTGTTTGCAAACCTTGACAGATGGTGC
 TGTGTTGATGCTAGCATTTCCAACGGTTTTTGGTGCAGCAGCAGCCACAGACTACACCTA
 CAATAAAGCTACAAATGATTTTCACTTTCAATGCCAGCATTGCTGCTGGTGTGCGGCCGG
 TGATAGTAACAGCGCAGCTCTGCAATCCTTCTGACTCCAAAAGCAGGTGATACAGCTAA
 CCTGAGCGTCAAAATCGGTACGACATCTGTTAATGTTGTTCTGGCGAGCGATGGCAAAAT
 TACAGCGAAAGATGGCTCAGCTCTGTATATCGACTCAACGGGTAACCTGACTCAGAACAG
 CGCAGGCACTGTAACAGCAGCAACCCTGGATGGACTGACCAAAAACCATGATGCGACAGG
 AGCTGTTGGTGTGATATCAGACCGCAGATGGCGCAACTATCTCTTGGCAGGCTCTGC
 TAACGCGGCAACAGGTACTCAATCAGGTGCAATTACACTGAAAAATGTTGCTATCAGTGC
 TGATGCTCTGCAGTCTGCTGCGAAAGGTACTGTTATCAATGTTGATAATGGTGTGATGA
 TATTTCTGTTAGTAAAACCGGGTGTGTTACTACCGGAGGTGCGCCTACTTATACTGATG
 CTGATGGTAAATTAACGACAACCAACACCGTTGATTATTTCTGCAAACTGATGGCAGCG
 TAACCAATGGTTCTGGTAAAGGGGTTTACACCGATGCAGCTGGTAAATTCATAACGACG
 CTGCAACCAAAGCCGCAACCACCACCGATCCGCTGAAAGCCCTTGATGACGCAATCAGCC
 AGATCGATAAGTTCCGTTTCATCCCTGGGTGCTATCCAGAACCGTCTGGATTCGCGGGTTA
 CCAACCTGAACAACACCACTACCAACCTGTCCGAAGCGCAGTCCCGTATTCAGGACGCCG
 ACTATGCGACCGAAGTGTCCAATATGTGAAAGCGCAGATCATCCAGCAGGCCGGTAACT
 CCGTGTGGCAAAAGCTAACCAAGGTACCGCAGCAGGTTCTGTCTGCTGCGGGTTAA

Figure 49

AACAAATCTCAGTCTTCTCTTAGCTCTGCTATTGAGCGTCTGTCTTCTGGT
 CTGCGTATTAACAGCGCAAAAGACGATGCAGCAGGTGAGGCGATTGCTAACCGTTTACG
 GCAAATATTAAAGGTCTGACCCAGGCTTCCCGTAACGCGAATGATGGTATTTCTGTTGCG
 CAGACCACTGAAGGTGCGCTGAATGAAATTAACAACAACCTGCAGCGTATTTCGTGAACTT
 TCTGTTTCAGGCAACTAACGGTACTAAGTCTGACAGCGATCTTCTTCTATCCAGGCTGAA
 ATTACTCAACGTCTGGAAGAAATTGACCGTGTATCTGAGCAAACCTCAGTTTAACGGCGTG
 AAAGTCCTTGCTGAAAATAATGAAATGAAAATTCAGGTTGGTGCTAATGATGGTGAAACC
 ATCACTATCAATCTGGCAAAAATTGATGCGAAAACCTCTCGGCCTGGACGGTTTTAATATC
 GATGGCGCGCAGAAAGCAACCGGCAGTGACCTGATTTCTAAATTTAAAGCGACAGGTACT
 GATAATTATCAAATTAACGGTACTGATACTATACTGTTAATGTAGATAGTGGAGTAGTA
 CAGGATAAAGATGGCAAAACAAGTTTATGTGAGTGCTGCGGATGGTTCACCTACGACCAGC
 AGTGATACTCAATTCAGATTGATGCAACTAAGCTTGCACTGGCTGCTAAAGATTTAGCT
 CAAGGTAATAAGATTGTCTACGAAGGTATCGAATTTACAAATACCGGCACTGGCGCTATA
 CCTGCCACAGGTAATGGTGAATTAACCGCCAATGTTGATGGTAAGGCTGTTGAATTCACCT
 ATTTCTGGGGAGTGCTGATACATCAGGTACTAGTGCAACCGTTGCCCTACGACAGCCCTA
 TACAAAATAGTGCAAGGCAATTGACTGCAACAAAAGTTGAAAATAAAGCAGCGACACTA
 TCTGATCTTGATCTGAACGCTGCCAAGAAAACAGGAAGCACGTTAGTTGTTAACGGTGCA
 ACTTACGATGTTAGTGAGATGGTAAAACGATAACGGAGACTGCTTCTGGTAACAATAAA
 GTCATGTATCTGAGCAAATCAGAAGGTGGTAGCCCGATTCTGGTAAACGAAGATGCAGCA
 AAATCGTTGCAATCTACCACCAACCGCTCGAACTATCGACAAAGCATTTGGCTAAAGTT
 GACAATCTGCGTTCTGACCTCGGTGCAGTACAAAACCGTTTCGACTCTGCCATCACCAC
 CTTGGCAACACCGTAACAACCTGTCTTCTGCCCCGTAGCCGTATCGAAGATGCTGACTAC
 GCGACCGAAGTGTCTAACATGTCTCGTGCGCAGATCCTGCAACAAGCGGGTACCTCTGTT CTGGCACAG

Figure 50

ATGGCACAAGTCATTAATACCAACAGCCTCTCGCTGATCAC
TCAAAATAATATCAACAAGAACCAGTCTGCGCTGTCGAGTTCTATCGAGCGTCTGTCTTC
TGGCTTGGCTATTAACAGCGCGAAGGATGACGCAGCGGGTCAGGCGATTGCTAACCGTTT
CACCTCTAACATTAAAGGCCTGACTCAGGCGGCCCCGTACGCCAACGACGGTATCTCCGT
TGCGCAGACCACCGAAGGCGCGCTGTCCGAAATCAACAACAACTTACAGCGTGTGCGTGA
ACTGACGGTACAGGCCACTACCGGTACTAACTCTGAGTCTGATCTGTCTTCTATCCAGGA
CGAAATTAAATCCCGTCTGGATGAAATTGACCGCGTATCTGGTCAGACCCAGTTCAACGG
CGTGAACGTGCTGGCAAAAAATGGCTCCATGAAAATCCAGGTTGGCGCAAATGATAACCA
GACTATCACTATCGATCTGAAGCAGATTGATGCTAAACTCTTGGCCTTGATGGTTTTAG
CGTTAAAAATAACGATACAGTTACCACTAGTGCTCCAGTAACTGCTTTTGGTGCTACCAC
CACAAACAATATTAAACTTACTGGAATTACCCCTTCTACGGAAGCAGCCACTGATACTGG
CGGAACTAACCAGCTTCAATTGAGGGTGTATTACTGATAATGGTAATGATTACTATGC
GAAAATCACCGGTGGTGATAACGATGGGAAGTATTACGCAGTAACAGTTGCTAATGATGG
TACAGTGACAATGGCGACTGGAGCAACGGCAAATGCAACTGTAAGTATGCAAACTACTAC
TAAAGCTACAACCTATCACTTCAGGCGGTACACCTGTTTCTGATTGATAATACTGCAGGTTT
CGCAACTGCCAACCTTGGTGCTGTTAGCTTAGTAAACTGCAGGATTTCAAGGGTAATGA
TACCGATACATATGCGCTTAAAGATACAAATGGCAATCTTTACGCTGCGGATGTGAATGA
AACTACTGGTGCTGTTTCTGTTAAACTATTACCTATACTGACTCTTCCGGTGCCGCCAG
TTCTCCAACCGCGGTCAAACCTGGGCGGAGATGATGGCAAAACAGAAGTGGTCGATATTGA
TGGTAAACATACGATTCTGCCGATTTAAATGGCGGTAACTCTGCAAACAGGTTTGACTGC
TGGTGGTGAGGCTCTGACTGCTGTTGCAAATGGTAAACACCGATCCGCTGAAAGCGCT
GGACGATGCTATCGCATCTGTAGACAAATCCGTTCTTCCCTCGGTGCGGTGCAAAACCG
TCTGGATTCCGCGGTTACCAACCTGAACAACACCACTACCAACCTGTCTGAAGCGCAGTC
CCGTATTCAGGACGCCGACTATGCGACCGAAGTGTCATATGTCGAAAGCGCAGATCAT
CCAGCAGGCCGGTAACTCCGTGTTGGCAAAAGCTAACCAGGTACCGCAGCAGGTTCTGTC
TCTGCTGCAGGGTTAA

Figure 51

ATGGCACAAGTCATTAATACCAACAGCCTCTCGCTGATCACT
 CAAAATAATATCAACAAGAACCAGTCTGCGCTGTCGAGTTCTATCGAGCGTCTGTCTTCT
 GGCTTGCGTATTAAACAGCGCGAAGGATGACGCCGAGGTCAGGCGATTGCTAACCGTTTT
 ACTTCTAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAACGACGGTATTCTGTGTT
 GCGCAGACCACCGAAGGCGCGCTGTCTGAAATCAACAACAACCTTACAGCGTATTCTGTGAA
 CTGACGGTTCAGGCTTCTACCGGGACTAACTCTGATTCGGATCTGGACTCCATTACAGGAC
 GAAATCAAATCCCGTCTGGACGAAATTGACCGCGTATCCGGTCAAACCCAGTTCAACGGT
 GTGAACGTACTGGCGAAAGACGGTTCGATGAAAATTACAGGTTGGTGCGAATGACGGCCAG
 ACTATCACTATTGATCTGAAGAAAATTGACTCTGATACGCTGGGGCTGAATGGTTTTAAC
 GTTAACGGCAAAGTACTATTGCGAACAAAGCGGCAACCATTAGTGATCTGGCGGCGACG
 GGGGCGAATGTTACTAACTCAAGCAATATTGTTGTACGACAAAGTTCAATGCCTTGAT
 GCAGCGACTGCATTTAGCAAACCTCAAAGATGGTGATTCTGTTGCCGTTGCTGCTCAGAAA
 TATACTTATAACGCATCGACCAATGATTTTACGACAGAAAATACAGTAGCGACAGGCACT
 GCAACGACAGATCTTGGCGCTACTCTGAAGGCTGCTGCTGGGCAGAGTCAATCAGGTACA
 TATACCTTTGCAAATGGTAAAGTTAACTTTGATGTTGATGCAAGCGGTAATATCACTATT
 GGCGGCGAAAAGGCTTTCTTGGTTGGTGGAGCGCTGACTACTAACGATCCCACCGGCTCC
 ACTCCAGCAACGATGTCTTCCCTGTTTAAGGCCGCGGATGACAAAGATGCCGCTCAATCC
 TCGATTGATTTTGGCGGGAAAAATACGAATTTGCTGGTGGCAATTCTACTAATGGTGGC
 GGCGTTAAATTCAAAGACACGGTGTCTTCTGACGCGCTTTTGGCTCAGGTTAAAGCGGAT
 AGTACTGCTAATAATGTAAAAATCACCTTTAACAATGGTCTCTGTCACTCACTGCATCG
 TTCCAAAATGGTGTATCTGGCTCCGCGGCATCGAATGCAGCCTACATTGATAGCGAAGGC
 GAACTGACAACTACTGAATCCTACAACACAAATTATTCCGTAGACAAAGACACGGGGGCT
 GTAAGTGTTACAGGGGGAGCGGTACGGGTAAATACGCCGCAAACGTGGGTGCTCAGGCT
 TATGTAGGTGCAGATGGTAAATTAACCACGAATACTACTAGTACCGGCTCTGCAACCAA
 GATCCACTAAATGCGCTGGATGAGGCAATTGCATCCATCGACAAATTCCGTTCTTCCCTG
 GGGGCTATCCAGAACCGTCTGGATTCCGCAGTCACCAACCTGAACAACACCACTACCAAC
 CTGTCTGAAGCGCAGTCCCGTATTACAGACGCCGACTATGCGACCGAAGTGTCCAACATG
 TCGAAAGCGCAGATCATCCAGCAGGCCGTAACCTCCGTGTTGGCAAAGCTAACCAGGTA
 CCGCAGCAGGTTCTGTCTCTGCTGCAGGGTTAA

Figure 52

AACAAGAACCAGTCTGCGCTGTGCGAGTTCTATCGAGCGTCTGTC
TTCTGGCTTGCGTATTAACAGCGCGAAGGATGACGCCGCGGGTCAGGCGATTGCTAACCG
TTTTACTTCTAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAACGACGGTATTTTC
TGTTGCGCAGACCACCGAAGGCGCGCTGTCCGAAATTAACAACAACCTTACAGCGTGTGCG
TGAGCTGACTGTTTCAGGCGACCACCGGTACTAAGTCTGAGTCTGACCTGTCTTCTATCCA
GGACGAAATCAAATCTCGCCTGGAAGAGATTGATCGTGTTCAGTCAGACTCAATTTAA
CGGCGTGAATGTTTTGGCTAAAGATGGGAAAATGAACATTACAGGTTGGGGCAAGTGATGG
ACAGACTATCACTATTGATCTGAAAAAGATCGATTCTACACTAAACCTCTCCAGTTT
TGATGCTACAACTTGGGCACCAAGTGTTAAAGATGGGGCCACCATCAATAAGCAAGTGGC
AGTAGATGCTGGCGACTTTAAAGATAAAGCTTCAGGATCGTTAGGTACCCTAAAATTAGT
TGAGAAAGACGTAAGTACTATGTAAATGACACTAAAAGTAGTAAGTACTACGATGCCGA
AGTAGATACTAGTAAGGGTGAAATTAAGTTCAACTCTACAAATGAAAGTGGAAGTACTCC
TACTGCAGCGACGGAAGTAACTACTGTTGGCCGCGATGTAAATTTGGATGCTTCTGCACT
TAAAGCCAACCAATCGCTTGTGCGTGATAAAGATAAAGCGGCAATGATGCTTATATCAT
TCAGACCAAAGATGTAACAATAATCAATCAACTTTCAATGCCGCTAATATCAGTGATGC
TGGTGTCTTTATCTATTGGTGCATCTACAACCGCGCCAAGCAATTTAACAGCTGACCCGCT
TAAGGCTCTTGATGATGCAATTGCATCTGTTGATAAATTCGCTCTTCTCTCGGTGCCGT
TCAGAACCGTCTGGATTCTGCCATTGCCAACCTGAACAACCACTACCAACCTGTCTGA
AGCGCAGTCCCGTATTAGGACGCTGACTATGCGACCGAAGTGTCACATGTCGAAAGC
GCAGATTATCCAGCAGGCCGGTAACTCCGTGCTGGCAAAA

Figure 53

ATGGCACAAAGTCATTAATACCAACAGCCTCTCGCTGATCACTCAAAA
 TAATATCAACAAGAACCAGTCTGCGCTGTGCGAGTTCTATCGAGCGTCTGTCTTCTGGCTT
 GCGTATTAAACAGCGCGAAGGATGACGCAGCGGGTCAGGCGATTGCTAACC GTTTACCTC
 TAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCTAACGATGGTATCTCTCTGGCGCA
 GACCACTGAAGGCGCACTGTCTGAGATTAAACAACACTTACAACGTGTGCGTGAGTTGAC
 TGTACAGGCGACCACCGGTACTAACTCTGATTCTGACCTGGCTTCTATT CAGGACGAAAT
 CAAATCCCGTTTGTCTGAAATTGACCGGTATCCGGGCAGACCCAGTTCAACGGCGTGAA
 CGTATTGTCTAAAGATGGCTCCCTGAAAATTCAGGTTGGCGCAAATGATGGTCAGACTAT
 CTCTATCGACCTGAAGAAAATTGACTCTGATACTCTGGGTTTGAATGGTTTCAACGTAA
 TGGTTCTGGTACCATTGCAAACAAAGCGGCCACAATCAGTGACTTGACTGCTCAGAAAGC
 CGTTGACAACGGTAATGGTACTTATAAAGTTACAAC TAGCAACGCTGCACTTACTGCATC
 TCAGGCATTAAAGTAAGCTGAGTGATGGCGATACTGTAGATATTGCAACCTATGCTGGTGG
 TACAAGTTCAACAGTTAGTTATAAATACGACGCAGATGCAGGTAAC TT CAGTTATAACAA
 TACTGCAAAACAAAACAGTGCTGCGGCTGGAACTCTGGCAGATACTCTTCTCCCGGCAGC
 TGGCCAGACTAAAACCGGTACTTACAAGGCTGCTACTGGTGATGTTAACTTTAATGTTGA
 CGCAACTGGTAATCTGACAATTGGCGGACAGCAAGCCTACCTGACTACTGATGGTAACCT
 TACAACAAACAACTCCGGTGGTGCGGCTACTGCAACTCTTAAAGAGCTGTTTACTCTTGC
 TGGCGATGGTAAATCTCTGGGGAACGGCGGTACTGCTACCGTTACTCTGGATAATACTAC
 GTATAATTTCAAAGCTGCTGCGAACGTTACTGATGGTGCTGGTGT CATCGCTGCTGCTGG
 TGTAAC TTATACAGCCACTGTTTCTAAAGATGTCATTCTGGCACA ACTGCAATCTGCAAG
 TCAGGCAGCAGCAACCGCTACCGACGGTGATACTGTGCAACGATCAACTATAAATCTGG
 TGT CATGATCGGTTCCGCTACCTTTACCAATGGTAAAGGTACTGCCGATGGTATGACTTC
 TGGTACA ACTCCAGTCGTAGCTACAGGTGCTAAAGCTGTATATGTTGATGGCAACAATGA
 ACTGACTTCCACTGCATCTTACGATACGACTTACTCTGTCAACGCAGATACAGGCGCAGT
 AAAAGTGGTATCAGGTACTGGTACTGGTAAATTTGAAGCTGTTGCTGGTGCGGATGCTTA
 TGTAAGCAAAGATGGCAAATTAACGACAGAAACCACCA GTGCAGGCACTGCAACCAAAGA
 TCCTTTGGCTGCCCTGGATGCTGCTATCAGCTCCATCGACAAATTCGTTCTCTCCCTGGG
 TGCTATCCAGAACCGTCTGGATTCCGCAGTCACCAACCTGAACAACCACTACTAACCT
 GTCTGAAGCGCAGTCCCGTATTTCAGGACGCCGACTATGCGACCGAAGTG TCCAATATGTC
 GAAAGCGCAGATCATCCAGCAGGCCGGTAACTCTGTGTTGGCAAAAGCTAACCAGGTACC
 GCAGCAGGTTCTGTCTCTGCTGCAGGGTTAA

Figure 54

ATGGCACAAAGTCATTAATACCAACAGCC
 TCTCGCTGATCACTCAAAATAATATCAACAAGAACCAGTCTGCGCTGTCGAGTTCTATCG
 AGCGTCTGTCTTCTGGCTTGCGTATTAACAGCGCGAAGGATGACGCCGCAGGTCAGGCGA
 TTGCTAACCGTTTTACTTCTAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAACG
 ACGGTATTTCTGTTGCACAGACCACTGAAGGCGCGCTGTCCGAAATCAACAACAACCTTAC
 AGCGTATTCGTGAAGTACCGTTTACGGCTTCTACCGGGACTAACTCTGATTCGGATCTGG
 ACTCCATTACAGGACGAAATCAAATCCCGTCTCGACGAAATTGACCGCGTTTCCGGTCAGA
 CCCAGTTCAACGGCGTGAACGTGCTGGCGAAAGACGGTTCGATGAAGATTCAGGTTGGCG
 CGAATGACGGGCAGACCATCTCTATCGATTTGCAGAAAATTGATTCTTCAACGCTGGGAT
 TGAAAGGTTTCTCGGTATCAGGGAACGCATTAAAAGTTAGCGATGCGATAACTACAGTTC
 CTGGTGCTAATGCTGGCGATGCCCGGTTACGGTTAAATTTGGTGCGAACGATACCGCTG
 CTGCCGCAATGGCTAAAAATTGGAATAAGTGATACATCAGGCTTGTCCCTACATAACG
 TACAAAGCGCGGATGGTAAAGCGACAGGAACCTATGTTGTTCAATCTGGTAATGACTTCT
 ATTCGGCTTCCGTTAATGCTGGTGGCGTTGTTACGCTTAATACCACCAATGTTACTTTCA
 CTGATCCTGCGAACGGTGTTACCACAGCAACACAGACAGGTCAGCCTATCAAGGTCACGA
 CGAATAGTGCTGGCGCGGCTGTTGGCTATGTTACTATTCAAGGCAAAGATTACCTTGCTG
 GTGCAGACGGTAAGGATGCAATTGAAAACGGTGGTGACGCTGCAACAAATGAAGACACAA
 AAATCCAACCTACCGATGAACTCGATGTTGATGGTTCTGTAAAAACAGCGGCAACAGCAA
 CATTTTCTGGTACTGCAACCAACGATCCGCTGGCACTTTTAGACAAAGCTATCTCGCAAG
 TTGATACTTTCCGCTCCTCCCTCGGTGCCGTACAAAACCGTCTGGATTCTGCGGTCACCA
 ACCTGAATAACACCACCACCAACCTGTCTGAAGCGCAGTCCCGTATTACAGACGCCGACT
 ATGCGACCGAAGTGTTCAACATGTCGAAAGCGCAGATCATCCAGCAGGCGGGTAACTCTG
 TGCTGTCTAAAGCTAACCAGGTACCGCAGCAGGTTCTGTCTCTGCTGCAGGGTTAA

Figure 55

CTTCTCTAGCTCTGCTATTGAGCGTCTGTCTTCTGGTCTGCGTATTAACAGCGCAAAAG
 ACGATGCAGCAGGTGAGGCGATTGCTAACCGTTTTACGGCAAAATATTAAAGGTCTGACCC
 AGGCTTCCCCTAACCGCAATGATGGTATTCTGTTGCGCAGACCACTGAAGGTGCGCTGA
 ATGAAATTAACAACAACCTGCAGCGTATTCGTGAACTTTCTGTTGAGGCAACTAACGGTA
 CTAACCTGACAGCGATCTTTCTTCTATCCAGGCTGAAATTACTCAACGTCTGGAAGAAA
 TTGACCGTGTATCTGAGCAAACTCAGTTTAACGGCGTGAAAGTCCTTGCTGAAAATAATG
 AAATGAAAATTCAGGTTGGTGCTAATGATGGTGAAACCATTGACCTGCCCCACGATTAG
 ATACAACACTCAGTTAGTAACGTCGGAATCTTCATTCTCAGAATGACCCCTTCTCCAGCC
 CGCTGCAAAATTCAGACGGTGTCTGATAATTCAGCGTGGAGTGCGGGCGGCATTCTGTTATA
 ATCCTGCCGCCAGTCATTAATAATTTTCTGGCATGAACGATATCGCTGAACCAAGTGCTC
 ATTCAAACATTTCATCGCGAAATCGTCCGTTAAAGCTCTCAATAAATCCGTTCTGCGTTGG
 CTTGCCCGGCTGGATTAAGCGCAACTCAACACCATGCTCAAAGGCCATTGATCCAGTGC
 ACGGCAAGTGAACCTCCGGCCCTGGTCAGTTCTTATCGTCGCCGGATAGCCTCGAAACAG
 TGCAATGCTGTCCAGAATACGCGTGACCTGAACGCCGTGAAATCCCAAAGGCAACAGTGAC
 CGTCAGGCATTCTTTGTGAAATCATCGACGCAGGTGAAGACACTTGATCCTGCGACCGGT
 GGAAAGTGCGTCCATGACGAAATCCATCGACCAAGTCAGATTGGGCGCCGCCGACGGAG
 CAGCGGCAGACGTTCTGTTGCCAGCCCTTTACGACGCTTCTGCGTTTTACGCCACGGCC
 ACTGAGGTGATAAAGCCGGTACACGCGCTTATGATTAAACATGAAGCCCTTCACGGCGCAG
 CAACTGCCAAATACGACGGTAGCCAAAACGCCCTGCGTCCAGTGCCAGCTCAGTGATGCG
 CCCTGATAAATGCGCATCAGCAGCCGGACGGTGAGCCTCATAGCGGCAGGTGACAGGGA
 TAAACCTGTAAAGCTGCAGGCACGACGTTGCGACAGACCGGTGCGATCACACATCAACAT
 CACGGCTTCCCGCTTCTGGTCTGTGTCAGTACTTTGCCCCAAGAGCCACCTGAAGCGCC
 TCTTTATCCAGCTTGGCTTCGCGCAAGCAGCTTCTTGAGTCTGGTGTCTCTTCTCAAGC
 GACTTCAGGCGCTTAACCTTCAGGCACCTCCATACCGCCATACTTCTTACGCCAGGTGTAA
 AACGTGGCATCGGAAATGGCATGCTTGCGGCAGAGTTCACGGGCGGGTACCCAGCTTCG
 GCTTCGCGGAGAATACTGATGATCTGTTCTGTCGGAATAACGCTTCTTATGGGGATGTCC
 TCATGTGGCTTATGAAGACATTACTAACATCGGGGTGACTAATCAACGGGGAGCAGGTC
 ACCATCACTATCAATCTGGCAAAAATTGATGCGAAAATCTCGGCCTGGACGGTTTTAAT
 ATCGATGGCGCGCAGAAAGCAACCGGCAGTGACCTGATTTCTAAATTTAAAGCGACAGGT
 ACTGATAATTATCAAATTAACGGTACTGATAACTATACTGTTAATGTAGATAGTGAGTA
 GTACAGGATAAAGATGGCAAACAAGTTTATGTGAGTGCTGCGGATGGTTCACTTACGACC
 AGCAGTGATACTCAATTCAAGATTGATGCAACTAAGCTTGAGTGGCTGCTAAAGATTTA
 GCTCAAGGTAATAAGATTGTCTACGAAGGTATCGAATTTACAAATACCGGCACTGGCGCT
 ATACCTGCCACAGGTAATGGTAAATTAACCGCAATGTTGATGGTAAGGCTGTTGAATTC
 ACTATTTGCGGGAGTGCTGATACATCAGGTACTAGTGCAACCGTTGCCCCCTACGACAGCC
 CTATACAAAATAGTGCAGGGCAATTGACTGCAACAAAAGTTGAAAATAAAGCAGCGACA
 CTATCTGATCTTGATCTGAACGCTGCCAAGAAAACAGGAAGCACGTTAGTTGTTAACGGT
 GCAACTTACGATGTTAGTGAGATGGTAAAACGATAACGGAGACTGCTTCTGGTAACAAT
 AAAGTCATGTATCTGAGCAAATCAGAAGGTGGTAGCCCGATTCTGGTAAACGAAGATGCA
 GCAAAATCGTTGCAATCTACCACCAACCCGCTCGAAACTATCGACAAAGCATTTGGCTAAA
 GTTGACAATCTGCGTTCTGACCTCGGTGCAGTACAAAACCGTTTCTGACTCTGCCATCACC
 AACCTTGGAACACCGTAAACAACCTGTCTTCTGCCCCGTAGCCGTATCGAAGATGCTGAC
 TACGCGACCGAAGTGTCTAACATGTCTCGTGCGCAGATCCTGCAACAAGCGGGTACCTCT
 GTTCTGGCACAGGCTAACC

Figure 56

AACAAAAACCAGTCTGCGCTGTCGACTTCTATCGAGCGCCTCTCT
TCTGGTCTGCGCATTAAACAGCGCTAAAGATGACGCTGCGGGCCAGGCGATTGCTAACCGC
TTCACCTTCTAACATCAAAGGTCTGACTCAGGCCGCACGTAACGCCAACGACGGTATCTCT
CTGGCGCAGACCACTGAAGGCGCACTGTCTGAAATCAACAACAACCTTGACGCGTGTTCGT
GAACTGACCGTTCAGGCCACTACCGGTACTAACTCTGATTCTGACCTGTCTTCAATCCAG
GACGAAATCAAATCCCGTCTCGATGAAATTGACCGCGTATCCGGTCAGACTCAGTTCAAC
GGCGTGAACGTACTGGCAAAAGATGGCTCGATGAAAATTCAGGTCGGTGCAAATGATGGT
CAGACAATCAGCATTGATTTCAGAGAAGATTGATTCTTCTACTTTAGGGTTAAATGGTTTT
TCTGTTTTCCAAAAATGCAGTATCTGTTGGTGATGCTATTACTCAATTGCCTGGCGAGACG
GCAGCCGATGCACCAGTAACCATCAAGTTTGATGATTGAGTAAAAACTGATTTAAAACTG
ACCGATGCTTCAGGGTTAAGTCTGCATAACCTCAAAGATGAAAATGGTAATTTAACTAAC
CAGTATGTTGTACAGAAATGGCGGAAAACTTTACGCTGCTACAGTCGCTGCCAATGGTAAT
GTTACGCTGAACAAAGCAAATGTAACCTACAGCGATGTCGCAAACGGTATTGATACCGCA
ACGCAGTCAGGCCAGTTAGTTTCAGGTTGGTGACAGATTCTACCGGTACGCCAAAAGCATT
GTGCTGTCCAAGGTAAAAGCTTTGGCATTGATGACGCCCGCTTGAAGAATAACACTGGT
GATGCTACCGCTACTCCACCGGAACATCTGGGACAACAGTTGTGCGAGCGTCAATTCAT
CTGAGTACGGGCAAAAACCTCTGTAGACGCTGATGTAACGGCTTCCACTGAATTCACAGGT
GCTTCAACCAACGATCCACTGACTCTGCTGGACAAAGCTATCGCATCTGTTGATAAAATTC
CGTTCTTCTTTGGGGCGGTACAGAACCGTCTGAGCTCCGCTGTAACCAACCTGAACAAC
ACCACCACCAACCTGTCTGAAGCGCAGTCCCGTATTCAGGACGCCGACTATGCGACCGAA
GTGTCCAACATGTCGAAAGCGCAGATTATCCAGCAGGCAGGTAACCTCCGTGCTGTCCAAA

Figure 57

AACAAAAACCAGTCTGCGCTGTCGACTTCTATCGAACGCCTCTCTTCTGG
CCTGCGTATTAACAGTGCGAAAGATGACGCTGCCGGTCAGGCGATAGCTAACCGTTTCAC
CTCTAACATTAAAGGCCTGACTCAGGCTGCGCGTAACGCCAACGACGGTATTTCTCTGGC
GCAGACCACAGAAGGTGCGTTGTCTGAAATCAACAACAACCTTGCAACGTGTGCGTGAGTT
GACCGTTCAGGCGACGACCGGTACTAACTCTGATTCTGACCTGTCATCTATTAGGACGA
AATCAAATCCCGTCTGGATGAGATTGACCGTGTTCGCGTCAGACCCAGTTCAACGGCGT
GAATGTACTGGCAAAAGACGGTTCGATGAAGATTAGGTTGGCGCGAATGATGGCCAGAC
TATTAGCATTGATTTACAGAAAATTGACTCTTCTACATTAGGGTTGAATGGTTTCTCCGT
TTCTGCTCAATCACTTAACGTTGGTGATTCAATTACTCAAATTACAGGAGCCGCTGGGAC
AAAACCTGTTGGTGTGATTTCACTGCTGTTGCGAAAGATCTGACTACTGCGACAGGTAA
AACTGTCGATGTTTCCAGCCTGACGTTACACAACACCCTGGATGCGAAAGGGGCTGCCAC
CGCACAGTTCGTCGTTCAATCCGGTAGTGATTTCTACTCCGCGTCCATTGACCATGCAAG
TGGTGAAGTGACGTTGAATAAAGCCGATGTGGAATACAAAGACACCGATAATGGACTAAC
GACTGCAGCTACTCAGAAAGATCAGCTGATTAAAGTTGCCGCTGACTCTGACGGCGCGGC
TGCGGGATATGTAACATTCCAGGGTAAAACTACGCTACAACGGCTCCAGCGGCGCTTAA
TGATGACACTACGGCAACAGCCACAGCGAACAAGTTGTTGTTGAATTATCTACAGCAAC
TCCGACTGCGCAGTTCTCAGGGGCTTCTTCTGCTGATCCACTGGCACTTTTAGACAAAGC
CATTGCACAGGTTGATACTTTCCGCTCCTCCCTCGGTGCCGTTCAAACCGTCTGGACTC
TGCGGTAACCAACCTGAACAACACCACCACCAACCTGTCTGAAGCGCAGTCCCGTATTCA
GGACGCCGACTATGCGACCGAAGTGTCTAACATGTGAAAGCGCAGATCATCCAGCAGGC
GGGTAACCTGTGTGCTGTCTAAA

Figure 58

ATGGCACAAG TCATTAATAC CAACAGCCTC TCGCTGATCA CTCAAATAA TATCAACAAG
 AACCAGTCTG CGCTGTCGAG TTCTATCGAG CGTCTGTCTT CTGGCTTGCG TATTAACAGC
 GCGAAGGATG ACGCCGCGGG TCAGGCGATT GCTAACCGTT TTAATTCTAA CATTAAAGGC
 CTGACTCAGG CTGCACGTAA CGCCAACGAC GGTATTTCTG TTGCACAGAC CACTGAAGGC
 GCGCTGTCCG AAATCAACAA CAACTTACAG CGTATCCGTG AGCTGACGGT TCAGGCTTCT
 ACCGGGACTA ACTCTGATTC GGATCTGGAC TCCATTCAGG ACGAAATCAA ATCCCGTCTC
 GACGAAATTG ACCGCGTATC CGGTCAGACC CAGTTCAACG GCGTGAACGT ACTGGCAAAA
 GACGGTTCGA TGAAAATTCA GGTGTTGCG AATGACGGTG AAATATCAC TATCGACCTG
 AAGAAAATCG ATTCTGATAC TCTGGGTCTG AATGGTTTTA ACGTAAATGG TAAAGGTACT
 ATTACCAACA AAGCTGCAAC GGTAAGTGAT TTAATTCTG CTGGCGCGAA GTTAAACAC
 CACGACAGGT CTTTATGATC TGAAAACCGA AAATACCTTG TTAATACCG ATGCTGCATT
 CGATAAATTA GGGAAATGGCG ATAAAGTCAC CGTTGGCGGC GTAGATTATA CTTACAACGC
 TAAATCTGGT GATTTTACTA CCACCAAATC TACTGCTGGT ACGGGTGTAG ACGCCGCGGC
 GCAGGCTACT GATTCAGCTA AAAAACGTGA TGCCTTAGCT GCCACCCTTC ATGCTGATGT
 GGGTAAATCT GTTAATGGTT CTTACACCAC AAAAGATGGT ACTGTTTCTT TCGAAACGGA
 TTCAGCAGGT AATATACCA TCGGTGGAAG CCAGGCATAC GTAGACGATG CAGGCAACTT
 GACGACTAAC AACGCTGGTA GCGCAGCTAA AGCTGATATG AAAGCGCTGC TTAAAGCCGC
 GAGCGAAGGT AGTGACGGTG CCTCTCTGAC ATTCAATGGC ACTGAATATA CTATCGCAAA
 AGCAACTCCT GCGACAACCT CTCCAGTAGC TCCGTTAATC CCTGGTGGGA TTAATTATCA
 GGCTACAGTG AGTAAAGATG TAGTATTGAG CGAAACCAA GCGGCTGCCG CGACATCTTC
 AATTACCTTT AATTCCGGTG TACTGAGCAA AACTATTGGG TTTACCGCGG GTGAATCCAG
 TGATGCTGCG AAGTCTTATG TGGATGATAA AGGTGGTATT ACTAACGTTG CCGACTATAC
 AGTCTCTTAC AGCGTTAACA AGGATAACGG CTCTGTGACT GTTGCCGGGT ATGCTTCAGC
 GACTGATACC AATAAAGATT ATGCTCCAGC AATTGGTACT GCTGTAAATG TGAATCCGC
 GGGTAAATC ACTACTGAGA CTACCAGTGC TGGTTCTGCA ACGACCAACC CGCTTGCTGC
 CCTGGACGAC GCTATCAGCT CCATCGACAA ATTCCGTTCT TCCCTGGGTG CTATCCAGAA
 CCGTCTGGAT TCCGAGTCA CCAACCTGAA CAACACCACT ACCAACCTGT CTGAAGCGCA
 GTCCCGTATT CAGGACGCCG ACTATGCGAC CGAAGTGTC AACATGTCGA AAGCGCAGAT
 TATCCAGCAG GCCGGTAACT CCGTGCTGGC AAAAGCCAAC CAGGTACCGC AGCAGGTTCT
 GTCTCTGCTG CAGGGTTAA

Figure 59

ATGGCACAAG TCATTAATAC CAACAGCCTC TCGCTGATCA CTCAAATAA TATCAACAAG
 AACCAGTCTG CGCTGTCGAG TTCTATCGAG CGTCTGTCTT CTGGCTTGCG TATTAACAGC
 GCGAAGGATG ACGCCGAGG TCAGGCGATT GCTAACCGTT TTAATTCTAA CATTAAAGGC
 CTGACTCAGG CGGCCGTAA CGCCAACGAC GGTATTTCTG TTGCGCAGAC CACCGAAGGC
 GCGCTGTCCG AAATCAACAA CAACTTACAG CGTATTCGTG AACTGACGGT TCAGGCCACT
 ACAGGGACTA ACTCCGATTC TGACCTGGAC TCCATCCAGG ACGAAATCAA ATCTCGTCTT
 GATGAAATTG ACCGCGTATC CGGCCAGACC CAGTTCAACG GCGTGAACGT GCTGGCGAAA
 GACGGTTCAA TGAAAATTCA GGTGCGTGC AATGACGGCG AAACCATCAC GATCGACCTG
 AAAAAAATCG ATTCTGATAC TCTGGGTCTG AATGGCTTTA ACGTAAATGG TAAAGGTACT
 ATTACCAACA AAGCTGCAAC GGTAAGTGAT TTAACCTCTG CTGGCGCGAA GTTAAACAC
 CACGACAGGT CTTTATGATC TGAAAACCGA AAATACCTTG TTAACACCG ATGCTGCATT
 CGATAAATTA GGAATGGCG ATAAAGTCAC AGTTGGCGGC GTAGATTATA CTTACAACGC
 TAAATCTGGT GATTTTACTA CCACTAAATC TACTGCTGGT ACGGGTGTAG ACGCCGCGGC
 GCAGGCTGCT GATTCAGCTT CAAAACGTGA TGCCTTAGCT GCCACCCTTC ATGCTGATGT
 GGGTAAATCT GTTAATGGTT CTTACACCAC AAAAGATGGT ACTGTTTCTT TCGAAACGGA
 TTCAGCAGGT AATATCACCA TCGGTGGAAG CCAGGCATAC GTAGACGATG CAGGCAACTT
 GACGACTAAC AACGCTGGTA GCGCAGCTAA AGCTGATATG AAAGCGCTGC TCAAAGCAGC
 GAGCGAAGGT AGTGACGGTG CCTCTCTGAC ATTCAATGGC ACAGAATATA CCATCGCAAA
 AGCAACTCCT GCGACAACCA CTCCAGTAGC TCCGTTAATC CCTGGTGGGA TTACTTATCA
 GGCTACAGTG AGTAAAGATG TAGTATTGAG CGAAACCAA GCGGCTGCCG CGACATCTTC
 AATTACCTTT AATTCCGGTG TACTGAGCAA AACTATTGGG TTTACCGCGG GTGAATCCAG
 TGATGCTGCG AAGTCTTATG TGGATGATAA AGGTGGTATC ACTAACGTTG CCGACTATAC
 AGTCTCTTAC AGCGTTAACA AGGATAACCG CTCTGTGACT GTTGCCGGGT ATGCTTCAGC
 GACTGATACC AATAAAGATT ATGCTCCAGC AATTGGTACT GCTGTAAATG TGAATCCGC
 GGGTAAATC ACTACTGAGA CTACCAAGTG TGGTTCTGCA ACGACCAACC CGCTTGCTGC
 CCTGGACGAC GCAATCAGCT CCATCGACAA ATTCCGTTCT TCCCTGGGTG CTATCCAGAA
 CCGTCTGGAT TCCGAGTCA CCAACCTGAA CAACACCACT ACCAACCTGT CCGAAGCGCA
 GTCCCGTATT CAGGACGCCG ACTATGCGAC CGAAGTGTCC AACATGTCGA AAGCGCAGAT
 CATTGAGCAG GCCGGTAACT CCGTGCTGGC AAAAGCTAAC CAGGTACCGC AGCAGGTTCT
 GTCTCTGCTG CAGGGTTAA

Figure 60

ATGGCACAAG TCATTAATAC CAACAGCCTC TCGCTGATCA CTCAAAATAA TATCAACAAG
 AACCAGTCTG CGTGTGCGAG TTCTATCGAG CGTCTGTCTT CTGGCTTGCG TATTAACAGC
 GCGAAGGATG ACGCAGCGGG TCAGGCGATT GCTAACCGTT TTACTTCTAA CATTAAAGGC
 CTGACTCAGG CTGCACGTAA CGCCAACGAC GGTATTTCTG TTGCGCAGAC CACCGAAGGC
 GCGCTGTCCG AAATCAACAA CAACTTACAG CGTATTCGTG AACTGACGGT TCAGGCCACT
 ACAGGGACTA ACTCCGATTC TGACCTGGAC TCCATCCAGG ACGAAATCAA ATCTCGTCTT
 GATGAAATTG ACCGCGTATC CGGCCAGACC CAGTTCAACG GCGTGAACGT GCTGGCGAAA
 GACGGTTCAA TGAAAATTCA GGTGCGTGCG AATGACGGCG AAACCATCAC GATCGACCTG
 AAAAAAATCG ATTCTGATAC TCTGGGTCTG AATGGCTTTA ACGTAAATGG TAAAGGTACT
 ATTACCAACA AAGCTGCAAC GGTAAGTGAT TTAACCTCTG CTGGCGCGAA GTTAAACAC
 CACGACAGGT CTTTATGATC TGAAAACCGA AAATACCTTG TTAACACCG ATGCTGCATT
 CGATAAATTA GGGAAATGGCG ATAAAGTCAC AGTTGGCGGC GTAGATTATA CTTACAACGC
 TAAATCTGGT GATTTTACTA CCACTAAATC TACTGCTGGT ACGGGTGTA ACGCCGCGGC
 GCAGGCTGCT GATTCAGCTT CAAAACGTGA TGCCTTAGCT GCCACCCTC ATGCTGATGT
 GGGTAAATCT GTTAATGGTT CTTACACCAC AAAAGATGGT ACTGTTTCTT TCGAAACGGA
 TTCAGCAGGT AATATCACC ACGGTGGAAG CCAGGCATAC GTAGACGATG CAGGCAACTT
 GACGACTAAC AACGCTGGTA GCGCAGCTAA AGCTGATATG AAAGCGCTGC TCAAAGCAGC
 GAGCGAAGGT AGTGACGGTG CCTCTCTGAC ATTCAATGGC ACAGAATATA CCATCGCAAA
 AGCAACTCCT GCGACAACCA CTCCAGTAGC TCCGTTAATC CCTGGTGGGA TTACTTATCA
 GGCTACAGTG AGTAAAGATG TAGTATTGAG CGAAACCAAA GCGGCTGCCG CGACATCTTC
 AATTACCTTT AATTCCGGTG TACTGAGCAA AACTATTGGG TTTACCGCGG GTGAATCCAG
 TGATGCTGCG AAGTCTTATG TGGATGATAA AGGTGGTATC ACTAACGTTG CCGACTATAC
 AGTCTCTTAC AGCGTTAACA AGGATAACGG CTCTGTGACT GTTGCCGGGT ATGCTTCAGC
 GACTGATACC AATAAAGATT ATGCTCCAGC AATTGGCACT GCTGTAAATG TGAATCCGC
 GGGTAAATC ACTACTGAGA CTACCAAGTG TGGTTCTGCA ACGACCAACC CGCTTGCTGC
 CCTGGACGAC GCAATCAGCT CCATCGACAA ATTCCGTTCT TCCCTGGGTG CTATCCAGAA
 CCGTCTGGAT TCCGCGGTCA CCAACCTGAA CAACACCACT ACCAACCTGT CCGAAGCGCA
 GTCCCGTATT CAGGACGCCG ACTATGCGAC CGAAGTGTCC AACATGTCGA AAGCGCAGAT
 CATCCAGCAG GCCGGTAACT CCGTGCTGGC AAAAGCTAAC CAGGTACCGC AGCAGGTTCT
 GTCTCTGCTG CAGGGTTAA

Figure 61

ATGGCACAAG TCATTAATAC CAACAGCCTC TCGCTGATCA CTCAAAATAA TATCAACAAG
 AACCAGTCTG CGCTGTCGAG TTCTATCGAG CGTCTGTCTT CTGGCTTGCG TATTAACAGC
 GCGAAGGATG ACGCCGCGGG TCAGGCGATT GCTAACCGTT TTACTTCTAA CATTAAAGGC
 CTGACTCAGG CTGCACGTAA CGCCAACGAC GGTATTTCTG TTGCACAGAC CACTGAAGGC
 GCGCTGTCCG AAATCAACAA CAACTTACAG CGTATCCGTG AGCTGACGGT TCAGGCTTCT
 ACCGGGACTA ACTCTGATTC GGATCTGGAC TCCATTGAGG ACGAAATCAA ATCCCGTCTC
 GACGAAATTG ACCGCGTATC CGGTCAGACC CAGTTCAACG GCGTGAACGT ACTGGCAAAA
 GACGGTTCGA TGAAAATTCA GGTGTTGTCG AATGACGGTG AAATATCAC TATCGACCTG
 AAGAAAATCG ATTCTGATAC TCTGGGTCTG AATGGTTTTA ACGTAAATGG TAAAGGTACT
 ATTACCAACA AAGCTGCAAC GGTAAGTGAT TTAACCTCTG CTGGCGCGAA GTTAAACAC
 CACGACAGGT CTTTATGATC TGAAAACCGA AAATACCTTG TTAACCTACG ATGCTGCATT
 CGATAAATTA GGGAAATGGCG ATAAAGTCAC CGTTGGCGGC GTAGATTATA CTTACAACGC
 TAAATCTGGT GATTTTACTA CCACCAAATC TACTGCTGGT ACGGGTGTAG ACGCCGCGGC
 GCAGGCTACT GATTCAGCTA AAAAACGTGA TGCGTTAGCT GCCACCCTTC ATGCTGATGT
 GGGTAAATCT GTTAATGGTT CTTACACCAC AAAAGATGGT ACTGTTTCTT TCGAAACGGA
 TTCAGCAGGT AATATCACC AATATCAGG CCAGGCATAC GTAGACGATG CAGGCAACTT
 GACGACTAAC AACGCTGGTA GCGCAGCTAA AGCTGATATG AAAGCGCTGC TTAAAGCCGC
 GAGCGAAGGT AGTGACGGTG CCTCTCTGAC ATTCAATGGC ACTGAATATA CTATCGCAAA
 AGCAACTCCT GCGACAACCT CTCCAGTAGC TCCGTTAATC CCTGGTGGGA TTACTTATCA
 GGCTACAGTG AGTAAAGATG TAGTATTGAG CGAAACCAA GCGGCTGCCG CGACATCTTC
 AATTACCTTT AATTCCGGTG TACTGAGCAA AACTATTGGG TTTACCGCGG GTGAATCCAG
 TGATGCTGCG AAGTCTTATG TGGATGATAA AGGTGGTATT ACTAACGTTG CCGACTATAC
 AGTCTCTTAC AGCGTTAACA AGGATAACGG CTCTGTGACT GTTGCCGGGT ATGCTTCAGC
 GACTGATACC AATAAAGATT ATGCTCCAGC AATTGGTACT GCTGTAAATG TGAATCCGC
 GGGTAAATC ACTACTGAGA CTACCAAGTG TGGTCTGCA ACGACCAACC CGCTTGCTGC
 CCTGGACGAC GCTATCAGCT CCATCGACAA ATTCCGTTCT TCCCTGGGTG CTATCCAGAA
 CCGTCTGGAT TCCGAGTCA CCAACCTGAA CAACACCACT ACCAACCTGT CTGAAGCGCA
 GTCCCGTATT CAGGACGCCG ACTATGCGAC CGAAGTGTC AACATGTGCA AAGCGCAGAT
 TATCCAGCAG GCCGGTAACT CCGTGCTGGC AAAAGCCAAC CAGGTACCGC AGCAGGTTCT
 GTCTCTGCTG CAGGGTTAA

Figure 62

ATGGCACAAGTCATTAATACCAACAGCCTCTCGCTGATCACTCAAAATAATATCAACAAG
 AACCCAGTCTGCGCTGTCTGAGTTCTATCGAGCGTCTGTCTTCTGGCTTGCGTATTAAACAGC
 GCGAAGGATGACGCCGAGGTGAGGCGATTGCTAACCGTTTTACTTCTAACATTAAAGGC
 CTGACTCAGGCGGCCCGTAACGCCAACGACGGTATTTCTGTTGCGCAGACCACCGAAGGC
 GCGCTGTCCGAAATCAACAACAACCTTACAGCGTATTCGTGAACTGACGGTTCAGGCCACT
 ACAGGGACTAACTCCGATTCTGACCTGGACTCCATCCAGGACGAAATCAAATCTCGTCTT
 GATGAAATTGACCGGTATCCGGCCAGACCCAGTTCAACGGCGTGAACGTGCTGGCGAAA
 GACGGTTCATGAAAATTCAGGTTGGTGCGAATGACGGCGAAACCATCACGATCGACCTG
 AAAAAATCGATTCTGATACTCTGGGTCTGAATGGCTTTAACGTAAATGGTAAAGGTACT
 ATTACCAACAAGCTGCAACGGTAAGTGATTAACTTCTGCTGGCGCGAAGTTAAACACC
 ACGACAGGTCTTTATGATCTGAAAACCGAAAATACCTTGTTAACTACCGATGCTGCATTC
 GATAAATTAGGGAATGGCGATAAAGTCACAGTTGGCGGCGTAGATTATACTTACAACGCT
 AAATCTGGTGATTTTACTACCACTAAATCTACTGCTGGTACGGGTGTAGACGCCGCGGCG
 CAGGCTGCTGATTGAGCTTCAAAACGTGATGCGTTAGCTGCCACCCCTTCATGCTGATGTG
 GGTAAATCTGTTAATGGTCTTACACCACAAAAGATGGTACTGTTTCTTTGAAACGGAT
 TCAGCAGGTAATATCACCATCGGTGGAAGCCAGGCATACGTAGACGATGCAGGCAACTTG
 ACGACTAACACCGCTGGTAGCGCAGCTAAAGCTGATATGAAAGCGCTGCTCAAAGCAGCG
 AGCGAAGGTAGTGACGGTGCCCTCTCTGACATTCAATGGCACAGAATATACCATCGCAAAA
 GCAACTCCTGCGACAACCACTCCAGTAGCTCCGTTAATCCCTGGTGGGATTACTTATCAG
 GCTACAGTGAGTAAAGATGTAGTATTGAGCGAAACCAAAGCGGCTGCCGCGACATCTTCA
 ATTACCTTTAATTCGGGTGTAAGTACGCAAACTATTGGGTTTACCGCGGGTGAATCCAGT
 GATGCTGCGAAGTCTTATGTGGATGATAAAGGTGGTATCACTAACGTTGCCGACTATACA
 GTCTCTTACAGCGTTAAACAAGGATAACGGCTCTGTGACTGTTGCCGGGTATGCTTCAGCG
 ACTGATACCAATAAAGATTATGCTCCAGCAATTGGTACTGCTGTAAATGTGAACTCCGCG
 GGTAAAATCACTACTGAGACTACCACTGCTGGTCTGCAACGACCAACCCGCTTGCTGCC
 CTGGACGACGCAATCAGCTCCATCGACAAATTCGGTCTTCCCTGGGTGCTATCCAGAAC
 CGTCTGGATTCCGCGAGTCACCAACCTGAACAACCACTACCAACCTGTCCGAAGCGCAG
 TCCCGTATTACAGGACGCCGACTATGCGACCGAAGTGTCCAACATGTGAAAGCGCAGATC
 ATTCAGCAGGCCGGTAACTCCGTGCTGGCAAAAGCTAACCAGGTACCGCAGCAGGTTCTG
 TCTCTGCTGCAGGGTTAA

Figure 63

ATGGCACAAG TCATTAATAC CAACAGCCTC TCGCTGATCA CTCAAATAA TATCAACAAG
 AACCAGTCTG CGCTGTCGAG TTCTATCGAG CGTCTGTCTT CTGGCTTGCG TATTAACAGC
 GCGAAGGATG ACGCCGCGAG TCAGGCGATT GCTAACCGTT TTAATTCTAA CATTAAAGGC
 CTGACTCAGG CTGCACGTAA CGCCAACGAC GGTATTTCTG TTGCGCAGAC CACCGAAGGC
 GCGCTGTCCG AAATCAACAA CAACTTACAG CGTATTCGTG AACTGACGGT TCAGGCCACT
 ACAGGGACTA ACTCCGATTC TGACCTGGAC TCCATCCAGG ACGAAATCAA ATCTCGTCTT
 GATGAAATTG ACCGCGTATC CGGCCAGACC CAGTTCAACG GCGTGAACGT GCTGGCGAAA
 GACGGTTCAA TGAAAATTCA GGTGTTGTCG AATGACGGCG AAACCATCAC GATCGACCTG
 AAAAAATCG ATTCTGATAC TCTGGGTCTG AATGGCTTTA ACGTAAATGG TAAAGGTACT
 ATTACCAACA AAGCTGCAAC GGTAAAGTAT TTAACCTCTG CTGGCGCGAA GTTAAACAC
 CACGACAGGT CTTTATGATC TGAAAACCGA AAATACCTTG TTAACACCG ATGCTGCATT
 CGATAAATTA GGAATGGCG ATAAAGTCAC AGTTGGCGGC GTAGATTATA CTTACAACGC
 TAAATCTGGT GATTTTACTA CCACTAAATC TACTGCTGGT ACGGGTGTAG ACGCCGCGGC
 GCAGGCTGCT GATTCAGCTT CAAAACGTGA TGCCTTAGCT GCCACCCTTC ATGCTGATGT
 GGGTAAATCT GTTAATGGTT CTTACACCAC AAAAGATGGT ACTGTTTCTT TCGAAACGGA
 TTCAGCAGGT AATATCACC AATATCACC TCGGTGGAAG CCAGGCATAC GTAGACGATG CAGGCAACTT
 GACGACTAAC AACGCTGGTA GCGCAGCTAA AGCTGATATG AAAGCGCTGC TCAAAGCAGC
 GAGCGAAGGT AGTGACGGTG CCTCTCTGAC ATTCAATGGC ACAGAATATA CCATCGCAAA
 AGCAACTCCT GCGACAACCA CTCCAGTAGC TCCGTTAATC CCTGGTGGGA TTACTTATCA
 GGCTACAGTG AGTAAAGATG TAGTATTGAG CGAAACCAA GCGGCTGCCG CGACATCTTC
 AATTACCTTT AATTCGGTG TACTGAGCAA AACTATTGGG TTTACGCGG GTGAATCCAG
 TGATGCTGCG AAGTCTTATG TGGATGATAA AGGTGGTATC ACTAACGTTG CCGACTATAC
 AGTCTCTTAC AGCGTTAACA AGGATAACGG CTCTGTGACT GTTGCCGGGT ATGCTTCAGC
 GACTGATACC AATAAAGATT ATGCTCCAGC AATTGGCACT GCTGTAAATG TGAATCCGC
 GGGTAAATC ACTACTGAGA CTACCACTGC TGGTTCTGCA ACGACCAACC CGCTTGCTGC
 CCTGGACGAC GCAATCAGCT CCATCGACAA ATTCCGTTCT TCCCTGGGTG CTATCCAGAA
 CCGTCTGGAT TCCGCGGTCA CCAACCTGAA CAACACCACT ACCAACCTGT CCGAAGCGCA
 GTCCCGTATT CAGGACGCGG ACTATGCGAC CGAAGTGTC AACATGTCGA AAGCGCAGAT
 CATCCAGCAG GCCGGTAACT CCGTGCTGGC AAAAGCTAAC CAGGTACCGC AGCAGGTTCT
 GTCTCTGCTG CAGGGTTAA

Figure 64

ATGGCACAAAG TCATTAATAC CAACAGCCTC TCGCTGATCA CTCAAAATAA TATCAACAAG
 AACCAGTCTG CGCTGTCGAG TTCTATCGAG CGTCTGTCTT CTGGCTTGCG TATTAACAGC
 GCGAAGGATG ACGCCGCGGG TCAGGCGATT GCTAACCGTT TTACTTCTAA CATTAAAGGC
 CTGACTCAGG CTGCACGTAA CGCCAACGAC GGTATTTCTG TTGCACAGAC CACTGAAGGC
 GCGCTGTCCG AAATCAACAA CAACTTACAG CGTATCCGTG AGCTGACGGT TCAGGCTTCT
 ACCGGGACTA ACTCTGATTC GGATCTGGAC TCCATTCAGG ACGAAATCAA ATCCCGTCTC
 GACGAAATTG ACCGCGTATC CGGTCAGACC CAGTTCAACG GCGTGAACGT ACTGGCAAAA
 GACGGTTCGA TGAAAATTCA GGTGCGTGC AATGACGGTG AAATATCAC TATCGACCTG
 AAGAAAATCG ATTCTGATAC TCTGGGTCTG AATGGTTTFA ACGTAAATGG TAAAGGTACT
 ATTACCAACA AAGCTGCAAC GGTAAAGTGAT TTAACCTCTG CTGGCGCGAA GTAAACACC
 ACGACAGGT CTTTATGATC TGAAAACCGA AAATACCTTG TTAACCTACG ATGCTGCATT
 CGATAAATTA GGAATGGCG ATAAAGTCAC CGTTGGCGGC GTAGATTATA CTTACAACGC
 TAAATCTGGT GATTTTACTA CCACCAAATC TACTGCTGGT ACGGGTGTAG ACGCCGCGGC
 GCAGGCTACT GATTCAGCTA AAAACCGTGA TGCGTTAGCT GCCACCCTTC ATGCTGATGT
 GGGTAAATCT GTTAATGGTT CTTACACCAC AAAAGATGGT ACTGTTTCTT TCGAAACGGA
 TTCAGCAGGT AATATCACCA TCGGTGGAAG CCAGGCATAC GTAGACGATG CAGGCAACTT
 GACGACTAAC AACGCTGGTA GCGCAGCTAA AGCTGATATG AAAGCGCTGC TTAAAGCCGC
 GAGCGAAGGT AGTGACGGTG CCTCTCTGAC ATTCAATGGC ACTGAATATA CTATCGCAAA
 AGCAACTCCT GCGACAACCT CTCCAGTAGC TCCGTTAATC CCTGGTGGGA TTTCTTATCA
 GGCTACAGTG AGTAAAGATG TAGTATTGAG CGAAACCAA GCGGCTGCCG CGACATCTTC
 AATTACCTTT AATTCCGGTG TACTGAGCAA AACTATTGGG TTTACCGCGG GTGAATCCAG
 TGATGCTGCG AAGTCTTATG TGGATGATAA AGGTGGTATT ACTAACGTTG CCGACTATAC
 AGTCTCTTAC AGCGTTAACA AGGATAACGG CTCTGTGACT GTTGCCGGGT ATGCTTCAGC
 GACTGATACC AATAAAGATT ATGCTCCAGC AATTGGTACT GCTGTAAATG TGAATCCGC
 GGGTAAAATC ACTACTGAGA CTACCACTGC TGGTCTGCA ACGACCAACC CGCTTGCTGC
 CCTGGACGAC GCTATCAGCT CCATCGACAA ATTCCGTTCT TCCCTGGGTG CTATCCAGAA
 CCGTCTGGAT TCCGCAGTCA CCAACCTGAA CAACACCACT ACCAACCTGT CTGAAGCGCA
 GTCCCGTATT CAGGACGCCG ACTATGCGAC CGAAGTGTC AACATGTCGA AAGCGCAGAT
 TATCCAGCAG GCCGGTAACT CCGTGCTGGC AAAAGCCAAC CAGGTACCGC AGCAGGTTCT
 GTCTCTGCTG CAGGGTTAA

Figure 65

ATGGCACAAG TCATTAATAC CAACAGCCTC TCGCTGATCA CTCAAAATAA TATCAACAAG
 AACCAGTCTG CGCTGTCGAG TTCTATCGAG CGTCTGTCTT CTGGCTTGCG TATTAACAGC
 GCGAAGGATG ACGCCGCAGG TCAGGCGATT GCTAACCGTT TTA CTCTCTAA CATTAAAGGC
 CTGACTCAGG CGGCCCGTAA CGCCAACGAC GGTATTTCTG TTGCGCAGAC CACCGAAGGC
 GCGCTGTCCG AAATCAACAA CAACTTACAG CGTATTCTGT AACTGACGGT TCAGGCCACT
 ACAGGGACTA ACTCCGATTC TGACCTGGAC TCCATCCAGG ACGAAATCAA ATCTCGTCTT
 GATGAAATTG ACCGCGTATC CGGCCAGACC CAGTTCAACG GCGTGAACGT GCTGGCGAAA
 GACGGTTCAA TGAAAATTCA GGTTGGTGCG AATGACGGCG AAACCATCAC GATCGACCTG
 AAAAAATCG ATTCTGATAC TCTGGGTCTG AATGGCTTTA ACGTAAATGG TAAAGGTACT
 ATTACCAACA AAGCTGCAAC GGTAAGTGAT TTA ACTTCTG CTGGCGCGAA GTTAAACAC
 CACGACAGGT CTTTATGATC TGAAAACCGA AAATACCTTG TTA ACTACCG ATGCTGCATT
 CGATAAATTA GGGAAATGGCG ATAAAGTCAC AGTTGGCGGC GTAGATTATA CTTACAACGC
 TAAATCTGGT GATTTTACTA CCACTAAATC TACTGCTGGT ACGGGTGTAG ACGCCGCGGC
 GCAGGCTGCT GATTCAGCTT CAAAACGTGA TGCGTTAGCT GCCACCCTTC ATGCTGATGT
 GGGTAAATCT GTTAATGGTT CTTACACCAC AAAAGATGGT ACTGTTTCTT TCGAAACGGA
 TTCAGCAGGT AATATCACCA TCGGTGGAAG CCAGGCATAC GTAGACGATG CAGGCAACTT
 GACGACTAAC AACGCTGGTA GCGCAGCTAA AGCTGATATG AAAGCGCTGC TCAAAGCAGC
 GAGCGAAGGT AGTGACGGTG CCTCTCTGAC ATTCAATGGC ACAGAATATA CCATCGCAAA
 AGCAACTCCT GCGACAACCA CTCCAGTAGC TCCGTTAATC CCTGGTGGGA TTACTTATCA
 GGCTACAGTG AGTAAAGATG TAGTATTGAG CGAAACCAAA GCGGCTGCCG CGACATCTTC
 AATTACCTTT AATTCGGTG TACTGAGCAA AACTATTGGG TTTACCGCGG GTGAATCCAG
 TGATGCTGCG AAGTCTTATG TGGATGATAA AGGTGGTATC ACTAACGTTG CCGACTATAC
 AGTCTCTTAC AGCGTTAACA AGGATAACGG CTCTGTGACT GTTGCCGGGT ATGCTTCAGC
 GACTGATACC AATAAAGATT ATGCTCCAGC AATTGGTACT GCTGTAAATG TGA ACTCCGC
 GGGTAAATC ACTACTGAGA CTACCAGTGC TGGTCTGCA ACGACCAACC CGCTTGCTGC
 CCTGGACGAC GCAATCAGCT CCATCGACAA ATTCCGTTCT TCCCTGGGTG CTATCCAGAA
 CCGTCTGGAT TCCGAGTCA CCAACCTGAA CAACACCACT ACCAACCTGT CCGAAGCGCA
 GTCCCGTATT CAGGACGCCG ACTATGCGAC CGAAGTGTC AACATGTCGA AAGCGCAGAT
 CATTGACGAG GCCGGTAACT CCGTGCTGGC AAAAGCTAAC CAGGTACCGC AGCAGGTTCT
 GTCTCTGCTG CAGGGTTAA

Figure 66

ATGGCACAAG TCATTAATAC CAACAGCCTC TCGCTGATCA CTCAAAATAA TATCAACAAG
 AACCAGTCTG CGCTGTCGAG TTCTATCGAG CGTCTGTCTT CTGGCTTGCG TATTAACAGC
 GCGAAGGATG ACGCCGCGGG TCAGGCGATT GCTAACCGTT TTAATTCTAA CATTAAAGGC
 CTGACTCAGG CTGCACGTAA CGCCAACGAC GGTATTTCTG TTGCACAGAC CACCGAAGGC
 GCGCTGTCTG AAATCAACAA CAACTTACAG CGTATCCGTG AGCTGACGGT TCAGGCTTCT
 ACCGGAAC TA ACTCTGATTC GGATCTGGAC TCCATTACAG ACGAAATCAA ATCCCGTCTT
 GATGAAATTG ACCGCGTATC CGGCCAGACC CAGTTCAACG GCGTGAACGT ACTGGCAAAA
 GACGGTTCGA TGAAAATTCA GGTGTTGTCG AATGACGGTG AAATATCAC TATCGACCTG
 AAGAAAATCG ATTCTGATAC TCTGGGTCTG AATGGTTTTA ACGTAAATGG TAAAGGTACT
 ATTACCAACA AAGCTGCAAC GGTAAGTGAT TTAACCTCTG CTGGCGCGAA GTTAAACAC
 CACGACAGGT CTTTATGATC TGAAAACCGA AAATACCTTG TTAATACCG ATGCTGCATT
 CGATAAATTA GGAATGCGC ATAAAGTCAC CGTTGGCGGC GTAGATTATA CTTACAACGC
 TAAATCTGGT GATTTTACTA CCACCAAATC TACTGCTGGT ACGGGTGTAG ACGCCGCGGC
 GCAGGCTACT GATTCAGCTA AAAAACGTGA TCGGTTAGCT GCCACCCTTC ATGCTGATGT
 GGGTAAATCT GTTAATGGTT CTTACACCAC AAAAGATGGT ACTGTTTCTT TCGAAACGGA
 TTCAGCAGGT AATATCACCA TCGGTGGAAG CCAGGCATAC GTAGACGATG CAGGCAACTT
 GACGACTAAC AACGCTGGTA GCGCAGCTAA AGCTGATATG AAAGCGCTGC TTAAAGCCGC
 GAGCGAAGGT AGTGACGGTG CTTCTCTGAC ATTCAATGGC ACTGAATATA CTATCGCAAA
 AGCAACTCCT GCGACAACCT CTCCAGTAGC TCCGTTAATC CCTGGTGGGA TTACTTATCA
 GGCTACAGTG AGTAAAGATG TAGTATTGAG CGAAACCAAA GCGGCTGCCG CGACATCTTC
 AATTACCTTT AATTCCGGTG TACTGAGCAA AACTATTGGG TTTACCGCGG GTGAATCCAG
 TGATGCTGCG AAGTCTTATG TGGATGATAA AGGTGGTATT ACTAACGTTG CCGACTATAC
 AGTCTCTTAC AGCGTTAACA AGGATAACGG CTCTGTGACT GTTGCCGGGT ATGCTTCAGC
 GACTGATACC AATAAAGATT ATGCTCCAGC AATTGGTACT GCTGTAAATG TGAATCCGC
 GGGTAAAATC ACTACTGAGA CTACCAGTGC TGGTTCTGCA ACGACCAACC CGCTTGCTGC
 CCTGGACGAC GCTATCAGCT CCATCGACAA ATTCCGTTCT TCCCTGGGTG CTATCCAGAA
 CCGTCTGGAT TCCGAGTCA CCAACCTGAA CAACACCACT ACCAACCTGT CTGAAGCGCA
 GTCCCGTATT CAGGACGCCG ACTATGCGAC CGAAGTGTC AACATGTGCA AAGCGCAGAT
 TATCCAGCAG GCCGGTAACT CCGTGCTGGC AAAAGCCAAC CAGGTACCGC AGCAGGTTCT
 GTCTCTGCTG CAGGGTTAA

Figure 67

ATGGCACAAG TCATTAATAC CAACAGCCTC TCGCTGATCA CTCAAAATAA TATCAACAAG
 AACCAGTCTG CGTGTCGAG TTCTATCGAG CGTCTGTCTT CTGGCTTGCG TATTAACAGC
 GCGAAGGATG ACGCCGCGGG TCAGGCGATT GCTAACCGTT TTAATTCTAA CATTAAAGGC
 CTGACTCAGG CTGCACGTAA CGCCAACGAC GGTATTTCTG TTGCACAGAC CACTGAAGGC
 GCGCTGTCCG AAATCAACAA CAACTTACAG CGTATCCGTG AGCTGACGGT TCAGGCTTCT
 ACCGGGACTA ACTCTGATTG GGATCTGGAC TCCATTGAGG ACGAAATCAA ATCCCGTCTC
 GACGAAATTG ACCGCGTATC CGGTCAGACC CAGTTCAACG GCGTGAACGT ACTGGCAAAA
 GACGGTTCGA TGAAAATTCA GGTGCGTGC AATGACGGTG AAATATCAC TATCGACCTG
 AAGAAAATCG ATTCTGATAC TCTGGGTCTG AATGGTTTTA ACGTAAATGG TAAAGGTACT
 ATTACCAACA AAGCTGCAAC GGTAAGTGAT TTAATTCTG CTGGCGCGAA GTTAAACAC
 CACGACAGGT CTTTATGATC TGAAAACCGA AAATACCTTG TTAATACCG ATGCTGCATT
 CGATAAATTA GGAATGGCG ATAAAGTCAC CGTTGGCGGC GTAGATTATA CTTACAACGC
 TAAATCTGGT GATTTTACTA CCACCAAATC TACTGCTGGT ACGGGTGTAG ACGCCGCGGC
 GCAGGCTACT GATTCAGCTA AAAAACGTGA TCGGTTAGCT GCCACCCTTC ATGCTGATGT
 GGGTAAATCT GTTAATGGTT CTTACACCAC AAAAGATGGT ACTGTTTCTT TCGAAACGGA
 TTCAGCAGGT AATATCACCA TCGGTGGAAG CCAGGCATAC GTAGACGATG CAGGCAACTT
 GACGACTAAC AACGCTGGTA GCGCAGCTAA AGCTGATATG AAAGCGCTGC TTAAAGCCGC
 GAGCGAAGGT AGTGACGGTG CCTCTCTGAC ATTCAATGGC ACTGAATATA CTATCGCAA
 AGCAACTCCT GCGACAACCT CTCCAGTAGC TCCGTTAATC CCTGGTGGA TTTCTTATCA
 GGCTACAGTG AGTAAAGATG TAGTATTGAG CGAAACCAA GCGGCTGCCG CGACATCTTC
 AATTACCTTT AATTCCGGTG TACTGAGCAA AACTATTGGG TTTACGCGG GTGAATCCAG
 TGATGCTGCG AAGCTTATG TGGATGATAA AGGTGGTATT ACTAACGTTG CCGACTATAC
 AGTCTCTTAC AGCGTTAACA AGGATAACGG CTCTGTGACT GTTGCCGGGT ATGCTTCAGC
 GACTGATACC AATAAAGATT ATGCTCCAGC AATTGGTACT GCTGTAAATG TGAATCCGC
 GGGTAAATC ACTACTGAGA CTACCAGTGC TGGTTCTGCA ACGACCAACC CGCTTGCTGC
 CCTGGACGAC GCTATCAGCT CCATCGACAA ATTCCGTTCT TCCCTGGGTG CTATCCAGAA
 CCGTCTGGAT TCCGAGTCA CCAACCTGAA CAACACCACT ACCAACCTGT CTGAAGCGCA
 GTCCCGTATT CAGGACGCCG ACTATGCGAC CGAAGTGTC AACATGTCGA AAGCGCAGAT
 TATCCAGCAG GCCGGTAACT CCGTGCTGGC AAAAGCCAAC CAGGTACCGC AGCAGGTTCT
 GTCTCTGCTG CAGGGTTAA

Figure 68

This Page is inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ BLACK BORDERS
- ☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☒ BLURED OR ILLEGIBLE TEXT OR DRAWING
- ☒ SKEWED/SLANTED IMAGES
- ☒ COLORED OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REPERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images
problems checked, please do not report the
problems to the IFW Image Problem Mailbox**